The Cambridge Psychological Library

THE ESSENTIALS OF MENTAL MEASUREMENT

CAMBRIDGE
UNIVERSITY PRESS
LONDON BENTLEY HOUSE
NEW YORK, TORONTO, BOMBAY
CALCUTTA, MADRAS MACMILLAN
TOKYO MARUZEN COMPANY LTD

THE ESSENTIALS OF MENTAL MEASUREMENT

BY

WILLIAM BROWN

M A , M D (Oxon), D Sc , F R C P (Lond)

WILDE READER IN MENTAL PHILOSOPHY AND DIRECTOR OF THE INSTITUTE OF EXPERIMENTAL PSYCHOLOGY IN THE UNIVERSITY OF OXFORD

AND

GODFREY H. THOMSON

D Sc , D C L (Dunelm), Ph D (Strasburg)

ANDREW BELL PROFESSOR OF THE THEORY, HISTORY AND PRACTICE OF EDUCATION, EDINBURGH UNIVERSITY

Fourth Edition

CAMBRIDGE

AT THE UNIVERSITY PRESS

1940

By WILLIAM BROWN

PSYCHOLOGY AND THE SCIENCES (Editor and Contributor) Adam and Charles Black, Ltd 1924

MIND AND PERSONALITY. University of London Press, Ltd. 1926

SCIENCE AND PERSONALITY (Terry Lectures, Yale) Oxford University Press 1929

PSYCHOLOGY AND PSYCHOTHERAPY Edward Arnold and Co 4th Edition 1940

MIND, MEDICINE AND METAPHYSICS Oxford University Press 1936 2nd Impression 1938

PSYCHOLOGICAL METHODS OF HEALING University of London Press, Ltd 1938

WAR AND PEACE ESSAYS IN PSYCHOLOGICAL ANALYSIS Adam and Charles Black, Ltd 1939

By GODFREY H THOMSON

INSTINCT, INTELLIGENCE AND CHARACTER AN EDUCATIONAL PSYCHOLOGY George Allen and Unwin, Ltd 1924 2nd Edition 1932

THE NORTHUMBERLAND MENTAL TESTS No 1 AND No 2 George G Harrap & Co , Ltd 1922

HOW TO CALCULATE CORRELATIONS A NON-MATHEMATICAL BOOK OF INSTRUCTIONS George G Harrap & Co , Ltd 1924

A MODERN PHILOSOPHY OF EDUCATION George Allen and Unwin, Ltd. 1929 2nd Edition 1931

THE MORAY HOUSE TESTS OF INTELLIGENCE, ENGLISH AND ARITHMETIC. University of London Press, Ltd , annually

THE FACTORIAL ANALYSIS OF HUMAN ABILITY University of London Press, Ltd 1939

First Edition 1911
(*By W. Brown*)
Second Edition 1921
Third Edition 1925
Fourth Edition 1940

PRINTED IN GREAT BRITAIN

PREFACE

THE present edition of *The Essentials of Mental Measurement* contains four new chapters (reprints of recent papers by each of the authors), to indicate in some measure, however inadequately, what changes have taken place in the subject, and in their opinions, since 1925 So much has happened since then, in that province of experimental psychology to which the book refers, that a completely new work would be required to cover the ground, and in the circumstances of the present year such a new book is impossible.

A good many pages of the volume (especially chapters IX and X) are taken up with a critical discussion of Professor Spearman's *Theory of Two Factors* and Thomson's *Sampling Theory* The present attitude of each author (Brown and Thomson) to this question can be gathered from the new chapters, and in Thomson's case from chapters III and XVIII of his book *The Factorial Analysis of Human Ability* (London and Boston, 1939). Thomson has shown that the equations of the two theories can be transformed into one another, in either direction, by an orthogonal transformation. Brown has in co-operation with Stephenson assembled a large battery of tests of intelligence which conforms strictly (within sampling limits) to the requirements of the *Theory of Two Factors* That theory has however itself undergone important developments and extensions in two directions For the one direction, the last sentence in chapter XI, written in 1924, has proved to be prophetic.

"Our position is", we then wrote, "that until the evidence is more clear we shall continue to suspect that numerous and wide group factors are present" The presence of such group factors, in addition to g, is now universally recognised, and that mainly because of the work of Professor Spearman and his disciples in tracing and identifying them—although this school regards the number of *general* group factors as small.

Indeed, there is now a school of thought, led by Thurstone, which has entirely dethroned g and replaced its "monarchic" rule by an "oligarchy" of group factors. In this present-day controversy we both find ourselves tending to prefer Spearman's rather than Thurstone's factors. Brown

does so more decidedly and confidently Thomson awaits further evidence but in the meantime leans to the Spearman system mainly because g is such a useful coefficient For him, the *Sampling Theory*, which is not incompatible with factorial *description*, is still however the most probable *causal explanation* of the interrelationships In this connection the second direction of development of factor theory has however to be considered.

That is, that the recognition has grown more and more explicit that mathematically many different systems of factors can describe the facts of the statistical interrelationships of tests, and there is much less confidence shown in ascribing any degree of reality to them, other than that attaching to mathematical coefficients. The choice between different systems has, it would seem, to be made on grounds of psychological convenience or utility or the like, and cannot be made, at any rate solely, on mathematical grounds.

WILLIAM BROWN
GODFREY H. THOMSON

February 1940

CONTENTS

PART I. PSYCHOPHYSICS

PART II. CORRELATION

ERRATA

P 30, l 3. The formula should be ‘$\frac{7!}{5!\,2!}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^2$’

P. 31, l 3 from bottom For ‘p’ *read* ‘p^k’

P 45, l 10 For ‘negative’ *read* ‘positive’.

P. 132, l 24. After ‘William Brown’s Formula’ *insert* ‘or the Spearman-Brown Formula’.

P. 167, l. 25. For ‘who found small trace of such order’ *read* ‘who found the evidence for such order inconclusive’.

PART I

PSYCHOPHYSICS

CHAPTER I

MENTAL MEASUREMENT

Equal appearing intervals—Just perceptible distances—The interpretation of Weber's Law—Indirect methods of measurement—The approach to measurement by means of grading magnitudes and their differences.

(1) EQUAL APPEARING INTERVALS

THE pre-conditions* of measurement in any sphere of experience are (1) the *homogeneity* of the phenomena, or of any particular aspect of it, to be measured, (2) the possibility of fixing a *unit* in terms of which the measurement may be made, and of which the total magnitude may be regarded as a mere multiple or sub-multiple. These pre-requisites are satisfied in the cases of spatial and temporal magnitudes, in terms of which, directly or indirectly, all the measurements of the physical sciences are expressed. It was thought by Fechner that they are also satisfied in the case of the strictly psychical phenomena of sensation-intensity, i e it was assumed that any given sensation-intensity might be regarded as made up of a sum of unit sensation-intensities. This view has been definitely rejected by many later psychologists in whose opinion every sensation-intensity is qualitatively distinct from every other sensation-intensity. "To introspection, our feeling of pink is surely not a portion of our feeling of scarlet, nor does the light of an electric arc seem to contain that of a tallow-candle in itself" (James)† Such writers contend that Fechner's mistake was due to a confusion of

* These pre-conditions are those usually stated But the idea of measurement has been so expanded during recent generations by the mathematical ideas of continuity, infinity and limit that they are becoming inadequate as a statement of the position. Compare the last section of the present chapter, on the approach to measurement by means of grading magnitudes and their differences.

† *Principles of Psychology*, I. p. 546.

sensation-intensities with the (physical) stimulus-values required to produce them.

Nevertheless, purely psychical measurement is not entirely impossible. Within any one series of sensation-intensities, e.g. a series of greys, the contrasts or "distances" separating different pairs of intensities are perfectly homogeneous with one another and can be measured in terms of one another or in terms of an arbitrarily chosen unit of "sense-distance." Given two brightness-intensities a and b, it is quite possible to find, within limits of error, a brightness-intensity c which is as much higher than b in the scale of intensities as b is than a, i.e. such that the sense-distance $\overline{bc}$ = the sense-distance $\overline{ab}$, or, again, it is quite possible, theoretically, to find a brightness-intensity d which bisects the sense-distance $\overline{ab}$, i.e. which is such that it is as far removed from a in the scale of intensities as b is from it—in symbols, $\overline{ad} = \overline{db}$. Hence the "distance," or disparity, of b from a is twice that of d from a, the distance of c from a is four times that of d from a. If, now, $\overline{ad}$, or the distance of d from a, be taken as a conventional unit, the values of $\overline{ab}$ and $\overline{ac}$ will be 2 and 4 respectively*.

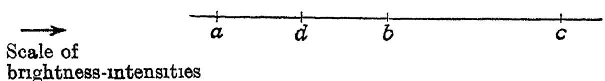


Fig. 1

A scale of intensities may in this way be formed rising by "equal-appearing intervals" or sense-distances, and the magnitude of any given interval may then, theoretically, be read off in terms of the unit-distance employed in the construction of the scale. In practice, however, it is found more convenient to fix the successive scale-marks, the successive members of the intensity-series, in terms of their corresponding stimulus-values. It has been found by experiment that in the case of light-intensities and sound-intensities the successive stimulus-values form, with fair approximation, a geometrical progression, or, in other words, each stimulus-value divided by the immediately preceding one gives approximately the same quotient. From the stimulus-values corresponding to an ascending series of eight equidistant brightness-values, Ebbinghaus obtained the following series of quotients:

2·3 2·1 2·1 1·8 1·7 1·7 2·0

* This view of mental measurement in terms of sense-distance first originated with J. R. L. Delboeuf, *Revue Philosophique*, 1878, v. p. 53. His term for sense distance was "contraste sensible.

The quotient value is not entirely constant, being slightly greater towards the two ends of the scale than it is at or about the middle. For this central region, then, the general relation of the sense-distance to the stimulus-values is given by the logarithmic formula

$$\text{Sense-distance } \overline{ab} = k \log \frac{\text{stimulus at } b}{\text{stimulus at } a},$$

where the stimulus at a is one which gives any finite intensity of sensation taken as the starting point or conventional zero (N B it is not necessarily liminal); the stimuli at a and b are those which correspond to the sensation-intensities of which $\overline{ab}$ is the "contraste sensible" or sense-difference.

(2) JUST PERCEPTIBLE DISTANCES

A mode of procedure which not only admits of much wider practical application than the above-mentioned "method of mean gradations," but also possesses a peculiar historical importance, is that which is concerned with the determination of the stimulus-increments corresponding to just-noticeable increments of sensation-intensity in different parts of the intensity scale. Weber found, in a series of experiments chiefly with lifted weights, that this stimulus-increment was relatively, not absolutely, constant for different regions of the intensity scale, i.e. that the stimulus corresponding to any original sensation-intensity had always to be increased by a constant proportion to arouse a just-noticeable increment of the sensation-intensity. If a 103 grams weight is just noticeably heavier than a 100 grams weight, then the weight just noticeably heavier than a 200 grams weight will be a 206 grams weight, not 203 grams

Mathematically formulated, Weber's Law is:

$$\frac{\delta\ (\text{stimulus})}{\text{stimulus}} = \text{a constant}.$$

The quantity δ (stimulus)/stimulus—$\frac{3}{100}$ or ·03 in our example of weight-lifting—is known as the "relative difference limen." It is of course the average value of a considerable number of determinations

Fechner verified Weber's Law in many different realms of sensation-intensity, and made it the basis of his own system of mental measurement. This he did by making the following three assumptions:

(1) that a sensation-intensity is a measurable magnitude and may therefore be regarded as a sum of unit-intensities;

(2) that just-noticeable differences of sensation-intensity are equal

at different parts of the stimulus scale, and may therefore conveniently serve as the unit-intensities above-mentioned,

(3) that the just-noticeable difference of sensation may be treated as a difference of two sensations, or at least that if Weber's Law applies to the former ("sensed difference") it will also apply to the latter ("difference sensation").

On the basis of Weber's Law and these added assumptions, Fechner obtains the following formula, viz.

$$d\,(\text{sensation}) = c\,\frac{d\,(\text{stimulus})}{\text{stimulus}},$$

which he calls the *fundamental formula for mental measurement.* Integrating, this becomes

$$\text{sensation} = c \log_e \text{stimulus} + C.$$

Putting the stimulus in this equation equal to the stimulus T for which the sensation is just below the threshold of consciousness, i.e. $= 0$, we have

$$0 = c \log_e T + C.$$

Subtracting the second from the first equation,

$$\text{sensation} = c \log_e \frac{\text{stimulus}}{T}.$$

Putting $T = 1$, and transferring to the ordinary logarithm system, we get

$$\text{sensation} = k \log \text{stimulus}.$$

All the assumptions involved are questionable. The first one has already been considered at some length. It is not the single sensation-intensity which is measurable, but the distinctness, disparity or distance of one sensation-intensity from another.

We must therefore regard the just-noticeable difference, not as a difference of two sensation-intensities but as a minimal sense-distance, if we are to be able to make use of it in our scheme of mental measurement. This modification, however, still leaves us involved in the difficulties of assumptions (2) and (3).

Fechner's own reason for regarding all just-noticeable differences belonging to any one scale of intensities as equal was that they appear equal to introspection. Introspection in a case like this is obviously difficult, even for the most skilful observers, and its verdict cannot, therefore, be greatly relied upon. Theoretically it is quite conceivable that just-noticeable differences, though equivalent to one another as being all just-noticeable, i.e. as being sense-distances so small that the

slightest diminution of them would cause them all, equally, to cease to be noticeable, yet as noticed or perceived would appear of different magnitude one from another. Ebbinghaus points to the analogous case of differentials or infinitesimals in mathematics These are all equivalent to one another as being all equally negligible as compared with finite magnitudes, yet are by no means necessarily equal to one another. If they belong to different "orders," those of a higher order are negligible as compared with those of a lower, etc Again, "the least distances perceived as such at different parts of the skin or in direct and indirect vision do not by any means all appear as equal magnitudes. On the contrary, so soon as they come to consciousness as distances they are at once perceived as distances of varying size, in a certain approximation to their objective differences*" In spite of these considerations, Ebbinghaus regards the correspondence of the stimulus results obtained for equal appearing intervals and for just perceptible intervals in the case of brightness-intensities (in the middle region of the scale both series of stimuli form a geometrical progression) as sufficient evidence for the approximate equality of the latter intervals. Muller and Wundt had previously advanced the same argument.

Several experimental investigations have been made with the express purpose of testing the relation of the methods of minimal change and mean gradations. Titchener† sums up the theoretical basis of such experiments concisely as follows: "There are in reality two possible ways of working (1) We might take a series of stimulus-values, corresponding to a series, say, of eight successive just-noticeable differences of sensation, and thereafter directly compare the two half-distances, of four just-noticeable differences each, and decide upon their equality or inequality. This would be a direct method of experiment." "Or (2) we might determine a few just-noticeable differences of sensation at different parts of the stimulus scale, in order to establish the constancy of the relative difference limen, and thereafter work with supraliminal differences, and decide whether the same constancy holds This would be an indirect method, it is the method indicated by the authors just cited [Muller, Wundt, Köhler and Tannery]....Either of these two methods would, presumably, take us to our goal. The experimental work would be exceedingly difficult. Liminal determinations are always and intrinsically difficult, and, further, the judgments passed upon just-noticeable differences and upon supraliminal differences are, even under the most

* Ebbinghaus, *Grundzüge der Psychologie*, 2nd ed., 1905, p 524.

† "Experimental Psychology," II. *Instructor's Manual*, p. lxxvii.

favourable conditions, the expressions of radically different mental attitudes"

It should be mentioned here that the principal rival to Fechner's hypothesis—the "difference hypothesis"—is that first formulated by Plateau, and generally known as the "quotient hypothesis." Plateau adopted the psychophysical formula

$$\text{sensation} = c\,(\text{stimulus})^k$$

on the basis of experiments by the method of mean gradations. This implies, in the place of Fechner's fundamental formula, the formula

$$\frac{\delta\,(\text{sensation})}{\text{sensation}} = k\,\frac{\delta\,(\text{stimulus})}{\text{stimulus}},$$

in other words, it assumes that just-noticeable differences are relatively, not absolutely, equal sensation-magnitudes. Although Plateau himself withdrew his formula later, the "quotient hypothesis" still remains as the rival of Fechner's "difference hypothesis"

To return to the experiments. Merkel (1888) worked with brightnesses, pressures, and noises, and found that the stimulus corresponding to the sensation bisecting a supraliminal sense-distance was the arithmetical mean of the stimuli corresponding to the two terminal sensations, which would seem to support the quotient hypothesis, though such an inference is not entirely free from objection. Angell (1892) worked with noise-intensities by the method of mean gradations, avoided certain sources of error present in Merkel's form of procedure, and obtained results supporting the difference hypothesis, the stimulus of the bisecting sensation being the geometrical, not the arithmetical, mean of the terminal stimuli.

A more thorough investigation was carried out by W. Ament* in 1900. He used a series of Marbe greys for the brightnesses and employed the direct method, for the noise-intensities he used a Fechner sound pendulum and worked principally by the indirect method. The result reached supported the quotient hypothesis. But Ament's work has not escaped criticism. A repetition of his experiments on brightness-intensities by Fròbes† has failed to confirm his results. Ebbinghaus also has found, in careful experiments with rotating sectors, that just-noticeable differences in different parts of the scale of brightnesses are equal to one another. On the whole, therefore, the balance of evidence seems to be in favour of the "difference hypothesis"

* W. Ament, "Ueber das Verhaltnis der ebenmerklichen zu den uebermerklichen Unterschieden bei Licht- und Schall-intensitaten," *Phil. Studien*, 1900, XVI. pp. 135 ff.

† *Zeitschrift fur Psychologie*, XXXVI. p. 344.

(3) THE INTERPRETATION OF WEBER'S LAW

Fechner's third assumption—see above, p. 4—brings us to the question of the interpretation of Weber's Law.

There are three general forms of interpretation:

(1) the psychophysical (Fechner),
(2) the physiological (Muller, Ebbinghaus, James, etc.),
(3) the psychological (Wundt).

According to (1), the logarithmic transition takes place in passing from the physiological changes in the sensory centres of the cerebral cortex to the corresponding sensation-intensities The chief objection to this view of Fechner's is that Weber's Law is not exact. This same consideration supports (2), or the physiological view, according to which the transition occurs either at the inception of the stimulus in the sense organ, or somewhere between the nerve endings and their central connections in the sensory areas of the cortex Experimental results obtained by Waller* and Steinach† are in favour of this view. Waller stimulated a frog's eye with light of different intensities, and found that the corresponding "negative variations" set up in the optic nerve varied in intensity as the logarithm of the stimulus-values (approx). Steinach obtained similar results on stimulating the skin of a frog's thigh with weights and noting the negative variation in the attached nerve If the negative variation may be assumed to be proportional to the intensity of the nerve-current passing along the nerve, these results point to the conclusion that the logarithmic transition occurs in the sense organ and its sensory nerve endings.

Ebbinghaus‡ has constructed a theory based upon the conception of varying degrees of dissociability of complex molecules to account for the law and also for the deviations from it towards the two extremes of the intensity scale

The psychological view, (3), of the law, held by Wundt, regards it as a special case of the general psychological "law of relativity" Stimulus, physiological process, and pure sensation-intensity increase in simple proportion to one another. The logarithmic transition occurs in passing from mere sensation and sensation difference to apperceived sensation and apperceived sensation difference, and the intensities are

* Waller, "Points relating to the Weber-Fechner Law," *Brain*, 1895, xviii p 200

† Steinach, "Elektromotorische Erscheinungen an Hautsinnesnerven bei adaquater Reizung," *Pfluger's Arch* 1896, Bd. 63, S 495

‡ Ebbinghaus, "Ueber den Grund der Abweichungen von dem Weber'schen Gesetz bei Lichtempfindungen," *Pfluger's Arch.* 1889, Bd. 54, S. 113

apperceived always in relation to one another In addition to the objection that it regards the sensation-intensities themselves, not their distances from one another, as measurable magnitudes, this view is also open to the criticism that it furnishes no explanation of the widely varying size of the relative difference limen in different sense-departments (e g DL for brightness-intensities $= \frac{1}{100}$, DL for sound-intensities $=$ about $\frac{1}{8}$); the physiological view sees in the varying structure of the different sense organs the adequate explanation of this Moreover, the psychological view has no completely satisfactory explanation to give of the deviations from Weber's Law so frequently met with.

These objections and difficulties make the view improbable, but by no means prove it to be impossible. It has the great merit of emphasising more definitely than was heretofore the case the importance of the more purely psychological factors in psychophysical experiments—in particular, it brings into prominence the distinction between mere disparity of sensation-intensities present simultaneously or in immediate succession in the same consciousness, and the perception of this disparity, the *discrimination* of the intensities one from another. In psycho-physical experiments the subject's consciousness is not limited to the mere sensational level.

In this connection the distinction, explained in Chapter III of the present book, on page 75, between two essentially different measures of a subject's fineness of discrimination, is not without importance, for those two measures, as far as experimental work goes, are both believed to obey Weber's Law. The one is the difference threshold spoken of in the present discussion, the other is there defined and named the interquartile range of the point of subjective equality: and the two quantities appear to differ in the level, sensational or perceptual, at which they stand. The extended idea of the *probability* of a judgment there defined throws this into prominence.

Another fact which would seem to have a very immediate bearing on the question of the interpretation of Weber's Law is that plants in their response to the stimulus of gravity (Geotropism) appear to obey that law*.

Although with continuous increase of stimulus-intensities the corresponding sensation-intensities rise in steps, each representing a just-noticeable difference, the psychophysical relation is really a strictly

* H. Fitting, *Jahrbuch f. wiss. Botanik*, Bd. XLI. 1905; F. Darwin, *New Phytologist*, 1906; James Small, *Annals of Botany*, XXXI. April 1917; James Small, *Proc. Roy. Soc.* B, XC. 1918.

continuous one, as becomes at once obvious if we consider a special case. The sensation-intensity aroused in lifting a weight of 100 grams is "indistinguishable," as we say, from that aroused by 102 grams, the sensation aroused by 102 grams is "indistinguishable" from that aroused by 104 grams; yet the sensation aroused by 100 grams is perceptibly different from that aroused by 104 grams Thus the sensation-intensity increases continuously, and the reason why this is not immediately apparent is probably to be looked for in the physiological mechanism of the psychophysical organism. The fact is that statements like the above as to two sensations being "distinguishable" or not lack precision in the absence of a definition of what is to be meant by distinguishable. To introspection almost all sensations are distinguishable from one another inasmuch as we seldom will agree that two are identical. Moreover, although a man will not (in the case of a subject of average sensitivity) give a majority of answers *heavier* in comparing 102 grams with 100 grams, yet the number he does give will, if the experiment is sufficiently carefully performed, be greater than he will give with 101 grams as the comparison weight (100 still being the standard). Although therefore he does not give, either with 101 or with 102 grams, a majority of answers *heavier* in comparing them with a standard of 100 grams, yet he does distinguish them from one another (if we take the result of a number of experiments), giving more *heavier* answers with the heavier weight*.

Delboeuf held that the limen has no psychological importance whatever If this is an extreme view, the importance which Fechner attributed to the limen is equally extreme in the other direction.

The absolute or stimulus limen is similar in kind to the difference limen, since consciousness is never empty of sensation-intensities when such a limen is being determined. Here again, Fechner's rigid distinction of the two was a fundamental error

(4) INDIRECT METHODS OF MEASUREMENT

The preceding account has probably sufficed to show that purely psychical measurement is a conceivable possibility. Its practical application however has been more detailed than extensive. A more generally useful method in quantitative psychology is that which measures the external, physical or physiological, causes and effects of mental process. The measurements are made in terms of the physical units of space and

* Compare Poincaré, *Science and Hypothesis*, Scott, 1905, p. 22; and G. H. Thomson, *British Association Reports*, 1913, paper under sub-section I.

time, yet they are not merely physical measurements, since they derive all their significance from the correlated psychical processes. They are indirect psychical measurements*. Measurements of reaction-times, memory, fatigue, illusions, etc. are all of this nature. Their varieties are innumerable, and are illustrated by the accounts in any good text-book of experimental psychology (Sanford, Titchener, Myers). In all cases full introspective accounts are essential, and when correlated with the measurements make the latter essentially psychical measurements. Measurements of limina, referred to in the previous section, are of the same nature. They are of some special importance as being measures of sensory acuity, etc.—aspects of the total mental ability of psycho-physical organisms. They figure prominently in many researches based on the use of "mental tests."

A method which makes a partial return to the more purely psychical form of measurement in terms of "distance" is the method of ranks or grades. Suppose we are considering the relative abilities of, say, 100 boys in English Composition We should find it difficult to mark their essays individually in terms of any constant unit but might find it possible to arrange them in order of merit, especially if we had sufficient time at our disposal to employ the method of "paired comparisons." According to the procedure of this latter method, the essays would be taken in pairs, quite at random, and the better essay of each pair would be given a "preference mark." This procedure would be repeated again and again until every essay had been compared with every other essay. The order of merit is then given by the number of preferences attaching to each essay. In this order, however, we cannot assume that the "ability-distance" from one boy to the next is a constant quantity. The boys near the extreme ends of the series will be farther removed from one another than the boys near the middle. We could only adjust for this if we knew the law of frequency-distribution for this kind of ability in this particular species of boy, and theoretically the determination of this distribution depends upon a prior fixing of the psychological unit, the unit of "ability-distance"; so that strictly the problem is insoluble. Since however under certain definite and indefinite conditions the form of distribution in a large number of biological and other cases of "physical" measurement has been found to be either Gaussian (normal) or differing from normal in ways described by Pearson's family of frequency curves, we might, with some probability of being near the truth, *assume* the normal form of distribution in the given case and so

* See Ebbinghaus, *Grundzüge der Psychologie*, 1905, pp. 75, 76.

obtain a quantitative measure for the ability of each particular boy*. A direct psychological determination and application of the (conventional) unit-distance is not, perhaps, an entirely impossible problem, and work in this direction may be expected and if achieved would certainly be much more scientific and psychological than the present method of measuring in terms of the external quantum of work done.

Finally, the interrelations of different mental abilities within any well-defined group of individuals situated within any definite environment may be determined by means of the technical method of "correlation" A correlation coefficient or other similar constant (e g correlation ratio) measures the *tendency* towards concomitant variation of two mental or other abilities within a group of individuals The result may be transferred to any single individual within the group as measuring the degree of probability of connection of the two abilities in the particular case The correlation between two abilities may be due to an actual direct relation of the abilities to one another, or, indirectly, to the influence of a common external environment upon them both. The first of these two cases is perhaps the more important, but the possibility of the second should not be lost sight of, and it also has a special interest of its own. The problems of correlation will be considered more fully in a later chapter

(5) THE APPROACH TO MEASUREMENT BY MEANS OF GRADING MAGNITUDES AND THEIR DIFFERENCES

Since the publication of the first edition of this book an important symposium bearing on the question of mental measurement has been held (in 1913). The exact problem submitted to the joint meeting of the Mind Association, the Aristotelian Society, and the British Psychological Society, was "Are the Intensity-differences of Sensation Quantitative?" Although many of the arguments of that discussion† are beyond the province of this book, it is of interest to note that there was a general consensus of opinion that sensation-intensities, and their differences, are at least "magnitudes" which can be graded, even if they be not "quantities" which can be measured. And following out further a suggestion contained in a quotation from Mr Bertrand Russell made by Professor Dawes Hicks in his contribution to the symposium, it may here be pointed out that grading leads, if the differences can also be graded, to something almost indistinguishable from, if indeed it

* This was done, e g. by Professor Pearson in his paper in *Biometrika*, 1907, v p. 105, and the example has been followed with success by other workers

† The papers are published in the *British Journ. of Psychol.* 1913, vi pp 137—189.

be not identical with, true measurement. For if we can arrange in order of magnitude a, b, c, ... and also their first differences $a - b$, $b - c$, $c - d$, . and the differences of these differences in turn, and so on, we can space out the original quantities a, b, c, ... as accurately as though we used a unit and measured them. To take a simple example, suppose five quantities a, b, c, d, e have really the measures 10, 16, 20, 31, 32. If an observer, ignorant of these measures, only knows the order of grading a, b, c, d, e of the quantities, he has already made a considerable advance even although he does not know the spacing, or the distances apart. If however he further can grade the differences, that is, if he knows that the greatest difference is that between d and c, and that the others follow in the order $b - a$, $c - b$, $e - d$, he has advanced further towards measurement, in the sense of accurate spacing, although there are still many spacings that will satisfy these gradings. Thus far he can almost always go in mental phenomena. And although it is practically difficult, there does not seem any theoretical difficulty about taking the next step, and grading the differences of the second order. In our example this grading is

$$\begin{aligned} &(d - c) - (b - a) \text{ or } \alpha, \text{ say,} \\ &(c - b) - (e - d) \text{ or } \beta, \\ &(b - a) - (c - b) \text{ or } \gamma: \end{aligned}$$

and the order of the *third* differences is

$$(\alpha - \beta), \quad (\beta - \gamma).$$

If now we could have all these gradings we could space out the original quantities very closely indeed to their true positions. This can be best seen by attempting to alter some one of the values while leaving all these gradings unaltered Make d, for example, 29 instead of 31 and although the order a, b, c, d, e is unchanged, and also the order of the first differences, that of the second differences is completely altered.

With an infinite number of quantities, and all the gradings of all their differences, we should, it would seem, arrive at an exact solution of the problem, so that grading and measurement are not perhaps so different in their nature as might at first be thought.

Indeed a case could well be made out for the thesis that the theoretical objections sometimes brought against mental measurement really hold in the last resort against all measurement, and prove too much: and that the real difference between mental measurement and physical measurement is simply that mental phenomena, being practically more difficult to handle, force on our notice the epistemological difficulties inherent in all measurement, whereas in physical measurement familiarity has bred contempt.

CHAPTER II

THE ELEMENTARY THEORY OF PROBABILITY

Some statistical terms—Arithmetical short-cuts—Measures of scatter—The fundamental theorem in probability—The binomial expansion—The normal curve of error—Fitting a normal curve to distribution data—The method of least squares

(1) SOME STATISTICAL TERMS

THE theory of probability was developed chiefly from two classes of material, (*a*) games of chance such as coin or dice throwing, roulette, etc., and (*b*) statistics, as they are called, that is such collections of quantitative information as the census, trade returns, insurance data and the like. It is easy to see that psychological experiments frequently resemble both of these classes. Any experiment, the result of which depends upon a human decision, has much in common with the throw of a die In both cases it often seems mere chance what the result is, although we believe that in both cases this is due only to our ignorance of the numerous factors at work*. And, since this is so, any scientific experiment on the actions and reactions of human beings must necessarily be repeated many times, until there accumulates a mass of quantitative information similar in many respects to a census return.

The mass of quantitative information thus accumulated is found upon examination to have certain peculiarities or properties For example consider the following case —The experiments, carried out by Professor Urban in 1906—7, were on lifted weights. A standard weight of 100 grams was compared, by lifting it, with weights of 84, 88, 92, 96, 100, 104 and 108 grams. The standard weight was always lifted first, and as the second unknown weight was lifted the judgment lighter, equal or heavier, was given†.

Suppose the following answers were obtained on one occasion:

108	grams,	*answer*	heavier
104	,,	,,	equal
100	,,	,,	heavier
96	,,	,,	lighter
92	,,	,,	equal
88	,,	,,	lighter
84	,,	,,	lighter

* "The uncertainty of my judgment is in many occurrences so equally balanced as I would willingly compromise it to the deciding of chance and of the dice." Florio's *Montaigne*.

† See "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," by F. M Urban, *Archiv für die gesamte Psychologie*, 1909, xv. p. 261.

In this series the lowest answer heavier is at 100 grams. Let this experiment be repeated 400 times, and in every series let the position of the lowest answer *heavier* be recorded. In a particular case the distribution of these *just perceptibly heavier* points was as follows

Grams	84	88	92	96	100	104	108
Frequency . ..	1	8	36	85	143	119	8

These particulars are shown in graphic form in the adjoining figure where the points have been joined by straight lines to make a polygon, which shows at once the chief peculiarities of such a collection of data,

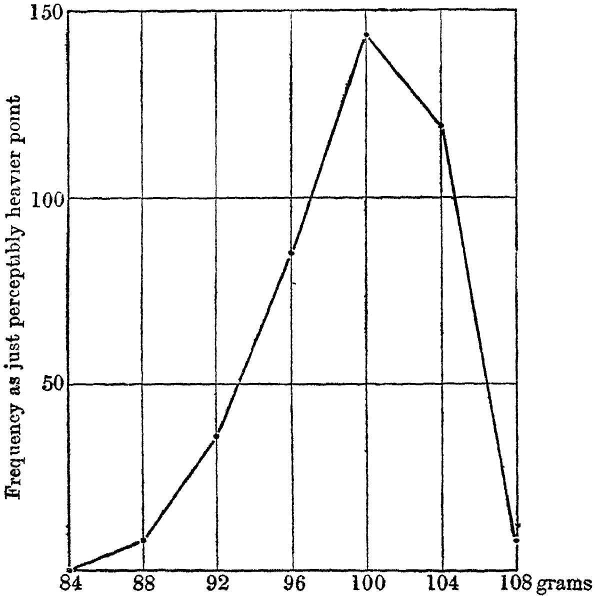


Fig. 2. A *cocked hat* curve (Urban's Subject III)

namely that the points are not scattered anyhow, but occur most frequently at a central value from which the frequency falls off in both directions Such a figure as this is colloquially termed a *cocked hat.* The point where the summit occurs, or point of greatest frequency, is called the *mode.* Here it is apparently at 100 grams, but if other weights between 100 and 104 grams could have been examined, it might be found to be elsewhere. The figure looks as though a smooth curve through the points would bring the mode somewhere a little higher than 100 grams.

The middlemost of the 400 positions of the just perceptibly heavier point is called the *median.* It too is in the 100 gram group, though here

again its exact position is uncertain. A better known central value (and one which can be more exactly calculated) is the *mean* or average value, found thus.

$$
\begin{array}{r}
1 \times 84 = 84 \\
8 \times 88 = 704 \\
36 \times 92 = 3312 \\
85 \times 96 = 8160 \\
143 \times 100 = 14300 \\
119 \times 104 = 12376 \\
8 \times 108 = 864 \\
\hline
\text{Divide by } 400\,)\,39800 \\
\hline
99\ 5 \text{ grams}
\end{array}
$$

In the above cases the only possible values of the sought for points were, by the nature of the experiments, at certain definite values In another class of experiment, however, all values within the range are possible as in the following case Twenty-nine experiments were made of bisecting a line by eye*. The lengths of the left-hand half of the line on these 29 occasions are given in this table, arranged in order of magnitude.

63 8 mms.	
62 2	
62·1	
61 4	
61 3	
61 2	
61 2	upper Quartile
61 1	
61 0	
60 9	
60 8	
60 6	
60 4	
60 3	
60 0	Median
59 9	
59 9	
59 6	
59 5	
59 2	
59 2	
59 1	lower Quartile
59 0	
58 8	
58 7	
58 6	
58 2	
58 1	
57 6	
29) 1743 7	
60 13 Mean.	

* An experiment performed for the purpose of illustrating this chapter Subject, G. H. Thomson

Here the median by counting is 60 mms. and the mean on calculation is found to be 60 13 mms. In this case the "cocked hat" is not at first sight evident, but if the data are grouped in some way it comes to light. For example, they can be arranged thus:

One reading	from	63	to	63 9	mms	inclusive
Two readings	,,	62	,,	62 9	,,	,,
Six ,,	,,	61	,,	61 9	,,	,,
Six ,,	,,	60	,,	60 9	,,	,,
Eight ,,	,,	59	,,	59 9	,,	,,
Five ,,	,,	58	,,	58 9	,,	,,
One reading	,,	57	,,	57 9	,,	,,

Here the numbers 1, 2, 6, 6, 8, 5, 1 show clearly the concentration in the middle, although being smaller they are not so regular as in the previous "cocked hat." Instead of concentrating these at points it seems more accurate to construct a diagram such as that here given

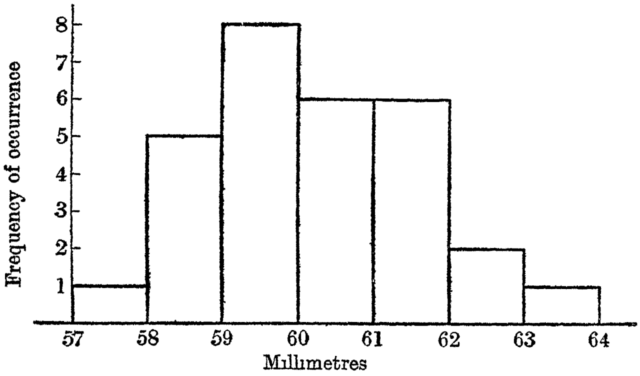


FIG. 3. A *histogram* formed from the bisection data

where a rectangle of the proper height is constructed over the corresponding base. Such a figure is called a "histogram."

This example is convenient for explaining two new terms, namely the "quartiles" and "semi-interquartile range." The "quartiles" are the readings which are one-quarter distant from each end, just as the "median" is half-way. In the table on page 15 there are 29 readings. The half-way reading is $\frac{29+1}{2}$ or 15 from either end, and the quartiles are at the readings $\frac{29+1}{4}$ or $7\frac{1}{2}$ from each end, for which we can take the mean of the seventh and eighth readings at 59·05 for the lower

quartile and 61·15 for the upper quartile. The semi-interquartile range is half the distance between the quartiles. It is, in this case,

$$\frac{61{\cdot}15 - 59{\cdot}05}{2} = 1{\cdot}05.$$

Clearly the semi-interquartile range is a crude measure of the scatter of the readings. When it is large the readings are more scattered, and therefore, any one of them is less likely to be correct, than if the semi-interquartile range had been small.

The mean value is sometimes called the *expectation* This term arises from games of chance For example, suppose I am playing against an opponent at a game of dice, in which my opponent has to give me as many shillings as there are pips in a single throw of the die, then the sum which I ought to give him before each throw, in order to make the game an even one, is three shillings and sixpence, which is my expectation of gain It is not the most probable sum for my opponent to give me, indeed he will never give me this exact sum, since he always pays me in shillings and not in pence But at the end of a sufficiently large number of throws I shall have received approximately this sum per throw on the average

From another point of view the mean will be found to correspond to a centre of gravity, namely the centroid of the curve of distribution of the readings. Let the adjoining figure represent such a curve, the number of readings which occur within a portion AB of the range being represented by the area included between the ordinates at A and B. Then the whole area of the curve will represent N the total number of readings. That is

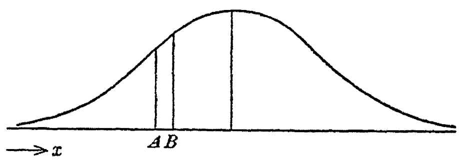


Fig. 4

$$N = \int y\,dx$$

where the integration is over the whole curve. Each reading is represented by its position on the x-axis, and the sum of all the readings is

$$\int xy\,dx.$$

The mean is therefore given by

$$\frac{\int xy\,dx}{\int y\,dx} \quad \text{or} \quad \frac{1}{N}\int xy\,dx.$$

But this is the expression for the abscissa of the centre of gravity of the curve, through which the ordinate above the mean therefore passes.

(2) SOME ARITHMETICAL SHORT-CUTS

There are several devices which considerably shorten the arithmetical labour of finding the mean. These can be illustrated by the example on page 15, near the top.

Choosing a convenient origin. The occurrence of large numbers in the calculations is minimised if an origin is chosen within the actual range of distribution, preferably either at one end or in the middle In the example we have in mind, the point 100 grams might be selected, as the large number 143 is thus avoided The calculations would then appear as follows.

1 × −16		−16
8 × −12		−96
36 × − 8		−288
85 × − 4		−340
		−740
119 × 4		476
8 × 8		64
		540
		−740
	400)	−200
		−0 5
	origin	100
		+99 5 Mean

Choosing a convenient unit Since in this particular case the measurements are all made at intervals of 4 grams, it is a considerable simplification if this is chosen as the unit. The same calculation then appears as follows:

1 × −4		−4
8 × −3		−24
36 × −2		−72
85 × −1		−85
		−185
119 × 1		119
8 × 2		16
		135
		−185
	400)	− 50
		− 125
		4 grams working unit
		− 5
	origin	100
		+99 5 Mean

The Summation Method of finding the Mean

This is only applicable in cases like the present where the readings are concentrated at a number of *equidistant* points. The calculation then appears as follows

84 grams	1	400
88 ,,	8	399
92 ,,	36	391
96 ,,	85	355
100 ,,	143	270
104 ,,	119	127
108 ,,	8	8
	400	) 1950
		4·875
		4 grams per unit
		19·5
	origin	80
		99·5 Mean

The figures in the second column are the continued sum, from below upwards, of the figures in the first column. They are then themselves added up, the total being 1950. Now it is clear from its method of formation that this total is composed as follows:

$$7 \times 8 + 6 \times 119 + 5 \times 143 + 4 \times 85 + 3 \times 36 + 2 \times 8 + 1 \times 1.$$

That is to say, it corresponds to taking the origin at 80 grams, and a unit of 4 grams.

It is instructive to try this calculation from the other end, as it were, making the continued sum from above downwards The reader should try other modifications of the idea underlying this summation method. For example, how should it be carried out in order to place the origin at 84 grams?

(3) MEASURES OF SCATTER

Let us consider again the experiment of bisecting a line, described on p. 15. Since the lines bisected measured 126 mms., the true half lay at 63 mms. The mean of the trials was 60·13 mms so that an error of 2·87 mms. was made. This error does not in itself however give a complete description of the subject's performance, a measure of scatter is required One such is the semi-interquartile range. When we say that the semi-interquartile range is 1·05 mms. we mean that half the attempts at bisection lay within a range of 2·1 mms., and, since half the trials already made come within this range, the probability of the next trial also coming within it is one-half, apart from practice improvement.

We must, however, next consider other and more usual ways of indicating scatter, although the use of the semi-interquartile range as a check is most valuable, *since it is a quantity about which glaring errors are not likely to be made.*

(1) There is first the "mean variation," a measure much in use among psychologists but not one to be recommended. In this the deviation or variation of each reading from the mean is written down, and the mean of these found, disregarding their sign. Taking the same example of bisecting a line we obtain

$$\text{M.V} = \frac{\text{Sum of deviations regardless of sign}}{\text{Number of cases } n} = 1{\cdot}14 \text{ mms.}$$

The mean variation is often, as here, about the same size as, or a little larger than, the semi-interquartile range, which latter, it will be seen, is indeed the median variation, from the median, though this name is not used

(2) The "standard deviation," generally denoted by σ, is by far the best measure of scatter to use, for reasons which will gradually become clear in the process of studying the subject. Its long title is the "root mean square deviation" and it is obtained by squaring each deviation, adding the squares all together, dividing by the number of readings and taking the square root.

$$\sigma = \sqrt{\frac{\text{Sum of squares of deviations}}{\text{Number of cases } n}}.$$

If this process be carried out the value $\sigma = 1{\cdot}38$ will be found for our example.

This is admittedly a longer process than finding the mean variation. It can be simplified by the following device, which should be used whenever the mean of the readings is a number involving awkward decimals. Instead of taking the deviations from the real mean, here 60·13 mms., let them be taken from some convenient point, here say 60 mms., chosen by mere inspection. Proceed now exactly as before, squaring and adding the deviations. But from the mean of the squared deviations must be deducted the square of the distance between the real mean and the convenient point from which the deviations have been taken (see example opposite).

The proof that this then gives the mean of the squares of the real deviations is easily obtainable by elementary algebra.

This shows one very important property of the mean, namely, that it is the point where the sum of the squares of the deviations is a

minimum· for the *provisional* mean square is always greater than the *real* mean square, since the correction subtracted is essentially positive.

Bisection data (page 15)

Deviations from 60 mms.	Squares of preceding
3 8	14 44
2 2	4 84
2 1	4 41
1 4	1 96
1 3	1 69
1 2	1 44
1 2	1 44
1 1	1 21
1 0	1 00
0 9	0 81
0 8	0 64
0 6	0 36
0 4	0 16
0 3	0 09
0 0	0 00
0 1	0 01
0 1	0 01
0 4	0 16
0 5	0 25
0 8	0 64
0 8	0 64
0 9	0 81
1 0	1 00
1 2	1 44
1 3	1 69
1 4	1 96
1 8	3 24
1 9	3 61
2 4	5 76

29) 55 71

1 92 provisional mean square

Subtract $0\ 13^2 = 0\ 02$

1 90 real mean square

$\sqrt{1\ 90} = 1\ 38 = \sigma.$

In this section we have, throughout, taken the *mean* as the central value from which the scatter was to be measured. In the great majority of cases this is the practical plan; but if necessary, measures of scatter from some other central value could be used. Indeed as has been pointed out, the semi-interquartile range is the median of the deviations from the *median*.

If the distribution of readings is represented by a smooth curve as

in the diagram on p 17, then it will be seen that the mean square deviation about the origin is given by the expression

$$\frac{1}{N}\int x^2 y\,dx$$

so that we have the equation, symbolising the above calculation,

$$\sigma^2 = \frac{1}{N}\int x^2 y\,dx - \text{Mean}^2,$$

the integrals being as before over the whole of the distribution.

In practical work, instead of actually giving the standard deviation, it is more usual to quote a quantity called the "probable error," which may, for the present, be arbitrarily defined as equal to $\cdot 67449\sigma$, i e. it is an arbitrary reduction of σ. To the meaning of this reduction we shall return presently.

The Standard Deviation about the True Value

In the section above we defined a quantity known as the Standard Deviation about the Mean. From another point of view, however, we sometimes require the value of the Standard Deviation about the True Value of the quantity which is being measured. This will always be a little larger than the former quantity, unless the true value happens to coincide exactly with the Mean. This follows from the theorem that the sum of the squares of the deviations is a minimum at the Mean.

Of course if we do not know the true value of what we are measuring, we shall be unable to find this quantity exactly, but it can be shown that it is approximated to if we divide by $n-1$ instead of n after finding the sum of the squares of the deviations

Let σ be standard deviation about mean a, and σ' standard deviation about true value A

Let $a_1, a_2, a_3, \ldots a_n$ be the readings.

Let $e_1, e_2, e_3, \ldots e_n$ be the true errors of those readings, and let e be the error of the mean a.

Then
$$a - a_1 = A + e - (A + e_1) = e - e_1,$$
$$a - a_2 = e - e_2, \text{ etc}$$

Therefore
$$\sigma^2 = \frac{S\,(a - a_k)^2}{n} = \frac{S\,(e - e_k)^2}{n} = \frac{ne^2 - 2eS\,(e_k) + S\,(e_k^2)}{n}.$$

Now
$$n\sigma'^2 = S\,(e_k^2)$$

and
$$ne = S\,(e_k),$$

so that we have

$$\sigma^2 = \sigma'^2 + e^2 - \frac{2eS\,(e_k)}{n} = \sigma'^2 + \frac{S^2\,(e_k)}{n^2} - \frac{2S^2\,(e_k)}{n^2}$$

$$= \sigma'^2 - \frac{S^2\,(e_k)}{n^2} = \sigma'^2 - \frac{S\,(e_k{}^2)}{n^2} - \frac{2S\,(e_k e_m)}{n^2}$$

The last term is approximately zero, the positive errors cancelling the negative.

We have finally

$$\sigma^2 = \sigma'^2 - \frac{S\,(e_k{}^2)}{n^2} = \sigma'^2 - \frac{\sigma'^2}{n},$$

$$\sigma = \sqrt{\frac{n-1}{n}}\,\sigma', \quad \therefore \sigma' = \sqrt{\frac{S\,(\text{deviation}^2)}{n-1}}.$$

It will be seen that when n is sufficiently large the two quantities become practically identical It is when n is small that the correction becomes of importance This is especially seen from a consideration of an extreme case, namely where only one measurement is made.

Let us say that the one measurement made has the value v Then the mean has also the value v, the deviation is zero, and the standard deviation is therefore the square root of zero divided by unity That is, the standard deviation about the mean is zero The standard deviation about the true value is, however, by the above rule, indeterminate, being the square root of zero divided by zero.

The reciprocal of the standard deviation is frequently used as a measure of the accuracy of a set of observations. It is clear from the above that if the number of observations is small it is the standard deviation about the true value which must be used, that is we must employ $n-1$ in the denominator instead of n. For otherwise the accuracy of a single measurement would be infinite.

The Standard Deviation of the Arithmetical Mean

In a section above it was shown that the mean square of the deviations of a set of readings is a minimum at the arithmetical mean. About any other value distant e from the mean this mean square is increased by e^2.

Now the mean square deviation about the mean is σ^2. But the mean square deviation about the true value is $n\sigma^2/(n-1)$. Therefore the expected square deviation of the mean from the true value is given by

$$e^2 = \frac{n\sigma^2}{n-1} - \sigma^2 = \frac{\sigma^2}{n-1}.$$

The expected square deviation (before experience) is, however, the same thing as the mean square deviation (after experience)*, so that the standard deviation of the mean about the *true value* is therefore

$$\frac{\sigma}{\sqrt{(n-1)}}$$

and that about the *mean of the means* will be

$$\sigma/\sqrt{n}.$$

The standard deviation of a mean is therefore obtained by dividing the standard deviation of the whole distribution by the square root of the number of readings.

The Standard Deviation of Sum or Difference

Let x be a quantity whose standard deviation is σ_x, and y a quantity whose standard deviation is σ_y; required the standard deviation of the sum $x+y$. Let m_x be the mean value of x and m_y the mean value of y. Then any single value of the sum $x+y$ will be of the form

$$m_x + \delta_x + m_y + \delta_y.$$

Moreover the mean of the sum $x+y$ is equal to $m_x + m_y$, so that the deviation of the above reading is

$$\delta_x + \delta_y.$$

The mean square deviation of the sum $x+y$ is therefore

$$\sigma^2_{x+y} = S\,(\delta_x + \delta_y)^2/n = S\,(\delta_x^2)/n + S\,(\delta_x\delta_y)/n + S\,(\delta_y^2)/n.$$

The quantity $S\,(\delta_x\delta_y)$ however will be very small if not zero, because the factors δ_x and δ_y are not connected with one another in any way, and their products are as likely to be positive as negative, so that the positive values will annul the negative on the average. We have, therefore, finally:

$$\begin{aligned} n\sigma^2_{x+y} &= S\,(\delta_x^2) + S\,(\delta_y^2) \\ &= n\sigma_x^2 + n\sigma_y^2, \\ \sigma_{x+y} &= \sqrt{(\sigma_x^2 + \sigma_y^2)}. \end{aligned}$$

Exactly the same reasoning holds for the value of σ_{x-y}, the only change being in the sign of m_y and δ_y, and therefore of $S\,(\delta_x\delta_y)$, which is however zero. The final result is the same

$$\sigma_{x-y} = \sqrt{(\sigma_x^2 + \sigma_y^2)}.$$

It is assumed in both these formulae that x and y are independent and uncorrelated.

* Contrast Keynes, *A Treatise on Probability*, London, 1921, pp 93 ff., on Venn's *Logic of Chance*.

(4) THE FUNDAMENTAL THEOREM IN PROBABILITY

In ordinary usage, when we say that an event is probable under certain circumstances, we mean that it is more likely to happen than not to happen, and by improbable we mean that it is more likely not to happen than to happen*

If we cross-question ourselves as to why we think an event is probable we often find that it is because we have more frequently found it to happen than to fail under similar circumstances in the past

If we wish to apply mathematical treatment to probability we must decide on a quantitative measure for it. We do so by using a fraction (vulgar or decimal) for this purpose, in such a way that the fraction rises and falls with the probability, becoming unity for "certain to happen" and zero for "certain not to happen." The numerator of this fraction is the number of equally probable ways in which the event can happen, while the denominator is the total number of equally probable ways in which the event can, under the given circumstances, either happen or fail†.

Thus, for example, consider the probability that with one throw of a six-faced die a score of more than four will be obtained. This can happen in two ways, namely, by throwing a five or a six. The total number of possible throws is six, and therefore the probability is $\frac{2}{6}$ or $\frac{1}{3}$.

This method of giving a quantitative value to a probability is clearly connected with the method adopted in betting‡ For instance, odds of "3 to 1 against" a certain event means that the speaker judges that there is only one chance of success to three chances of failure. The fraction representing the probability of success is therefore

$$\frac{1}{3+1}=\frac{1}{4}.$$

If the experiment with dice mentioned above be actually performed a large number of times, it will be found that the number of occasions on which the score exceeds 4 will closely approximate to one-third of the whole If therefore we had never seen dice and had no idea of their appearance, but were told that a large number of throws always included about one-third which were over four§, we should conclude that the probability of obtaining a throw of more than four was about one-third

* Cf Keynes, *op cit* ch VIII.

† Cf. however Jeffries and Wrinch, *Phil Mag* 1919.

‡ The odds at betting however depend also upon the existence of a market, cf Keynes, pp 22 and 23.

§ The late Professor Weldon found 106602/(12 × 26306) = 0 3377. See *Phil Mag.* June 1900, article by Professor K Pearson.

This method of finding probabilities, by deducing them from a large number of actual experiments, is that followed most frequently in practice.

For example if the points of an aesthesiometer are applied to a subject's forearm, with the points distant 3 cms. from one another, the average subject will usually recognise that two points are present. Occasionally however he will only feel one point. What is the probability that he will answer two? This can only be decided by experiment. In an actual case, 150 trials were made at a number of different sittings. On 105 occasions the answer two was returned If the conditions have been the same throughout then the probability of an answer two under these conditions is $\frac{105}{150} = 0\ 7$.

If the probability of an event occurring is p then the probability of it not occurring is clearly $1 - p$, for which q is often written. For the event is certain to happen or not to happen, and therefore the two probabilities must add up to "certainty" that is to unity. Thus in the above case the probability of an answer two *not* being returned is 0·3. Similarly the chance of throwing an ace at dice is $\frac{1}{6}$ and of not throwing an ace is $\frac{5}{6}$. The chance of obtaining a head on throwing a coin is $\frac{1}{2}$, and the chance of not obtaining a head is $\frac{1}{2}$

Next consider the probability of an event happening twice in succession, if its probability is p for one occurrence. Take a few specific cases first. If two coins are thrown or one coin thrown twice, there are four equally likely things than can happen, namely:

head	head
tail	tail
head	tail
tail	head

The probability of getting two heads is therefore $\frac{1}{4}$ which, it will be noticed, is equal to $(\frac{1}{2})^2$.

Next consider two throws of a six-faced die. There are here no less than 36 things which can happen, for whatever value the first throw has there are six different throws of the second die which can be associated with it: and since the first can also have six values there are 6×6 or 36 combinations possible. Only one of these combinations consists of two aces, so that the chance of throwing two aces is $\frac{1}{36}$ or $(\frac{1}{6})^2$. Generally, if the chance of an event is p then the chance of it occurring twice in succession is p^2. For suppose there are m ways in which it can happen, and n ways in which it can fail, then $p = m/(m + n)$. Also, having happened once, there are m ways in which the second success can occur,

each of which ways might be associated with the first success, and since there are also m ways of this first success happening there are in all m^2 ways of the double success happening. Similarly there are in all $(m+n)^2$ ways of the double event either happening or failing, and therefore the chance of a double success is

$$\frac{m^2}{(m+n)^2} = p^2.$$

Reasoning exactly similar to the above will enable the reader to convince himself of the truth of the following general theorem

If there are a number of *independent* events a, b, c, etc and their respective probabilities of occurrence are p_a, p_b, p_c, .. etc, then the probability of all occurring is the product

$$p_a \times p_b \times p_c \times \ldots$$

This may be said to be the fundamental proposition in probability. Its use will be recognised best by an example. It will be instructive to take a real psychometric experiment, that already described in the experiment on weight-lifting on p. 13 The subject of those experiments compared each of the weights with the standard 450 times. The results of this were as follows*:

Grams	84	88	92	96	100	104	108
No of answers heavier	0	11	32	100	212	402	431

From these the probability of the subject answering "heavier" at any weight can be calculated, always assuming that the conditions have throughout remained the same, by dividing the above numbers by 450 We thus obtain the following

Grams	84	88	92	96	100	104	108
Probability of answer "heavier"	0 0000	0 0244	0 0711	0 2222	0 4711	0 8933	0 9578

Let us now apply our theorem on the combination of probabilities to solve the following problem

Let the weights 84, 88 *grams etc. be each presented once to the subject What is the probability that the lowest answer "heavier" will be at* 96 *grams?*

The probability that he will answer "heavier" at 96 grams is 0·2222. The probability that he will *not* answer "heavier" at 92 grams is $1 - 0{\cdot}0711 = 0{\cdot}9289$ Similarly the probability that he will *not* answer "heavier" at 88 grams is 0·9756, and at 84 grams is 1·0000 The proba-

* Urban, *op cit* p 287, Table XI (multiply the values by 450) Or alternatively consult Table V, p. 175, of *The Application of Statistical Methods to the Problems of Psychophysics*, by F M Urban, Philadelphia, 1908

bility of the combination happening, namely, an answer "heavier" at 96 and none below, is, therefore

$$1{\cdot}0000 \times 0{\cdot}9756 \times 0{\cdot}9289 \times 0{\cdot}2222 = 0{\cdot}2014.$$

The actual frequency with which this occurred was 0·2125.

(5) IMPORTANCE OF THE BINOMIAL EXPANSION IN THE THEORY OF PROBABILITY

By a not unnatural hypothesis, which has been widely accepted and has proved very fruitful, an error of measurement, such as that made in judging the half-way point in a line in an experiment above, may be looked upon as the resultant of a large number of small circumstances, each of which sometimes sways our measurement in the one direction, sometimes in the other. This hypothesis, combined with the fundamental theorem of probability just explained, leads to the use of the binomial expansion in describing distributions of error. This can be illustrated by the following example —Let us suppose that a quantity which we desire to measure has really the value $13\frac{1}{2}$ units, but that we are opposed in our efforts to measure it by seven "Djinns," each of whom has the power of displacing our measurement by one half-unit. Let us further imagine that each of these mischievous imps, in an endeavour to prevent our making any steady measurement, decides that he will add or deduct his half-unit according to the throw, heads or tails, of a coin. Whenever we try to make a measurement, therefore, these invisible seven will assemble and throw each his coin in the air. If all the coins happen to come heads, seven half-units are cunningly added to the $13\frac{1}{2}$, and we obtain a measurement of 17. If all come tails, we get $13\frac{1}{2}$ minus $3\frac{1}{2}$, or 10. If on another occasion five are heads and two are tails, five half-units will be added and two subtracted, giving

$$13\tfrac{1}{2} + \tfrac{5}{2} - \tfrac{2}{2} = 15.$$

An actual test of this is given in the following figures and diagram. Seven coins were thrown on 128 occasions, and each time the proper number of half-units was added (for the heads) or subtracted (for the tails) from $13\frac{1}{2}$. The result was as follows:

0	heads	and	7	tails	occurred	2	times	giving	a	value	10 units
1	,,	,,	6	,,	,,	8	,,	,,	,,	,,	11 ,,
2	,,	,,	5	,,	,,	16	,,	,,	,,	,,	12 ,,
3	,,	,,	4	,,	,,	38	,,	,,	,,	,,	13 ,,
4	,,	,,	3	,,	,,	43	,,	,,	,,	,,	14 ,,
5	,,	,,	2	,,	,,	16	,,	,,	,,	,,	15 ,,
6	,,	,,	1	,,	,,	2	,,	,,	,,	,,	16 ,,
7	,,	,,	0	,,	,,	3	,,	,,	,,	,,	17 ,,
					Total	128					

This distribution is shown in Fig. 5.

The general resemblance between the diagram, made by throwing coins, and previous diagrams which represent the result of psychological experiments is not surprising if we consider for a moment what the condition of such experiments are. The "Djinns" which oppose our efforts to obtain a true value for say the spatial threshold are innumerable. Some are in the fingers of the experimenter, and make him press irregularly on the aesthesiometer points. Others cause noises to happen in the neighbourhood to distract the subject's attention. Other Djinns make the instrument hot one day and cold the next, others live in the subject's skin, and quite a lot are engaged in stirring up vivid imaginations in his mind so that he feels all kinds of prickles and

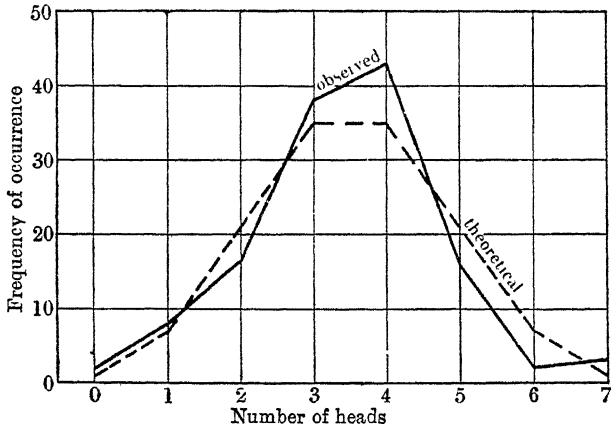


Fig 5 Number of heads in a throw of seven coins in 128 repetitions

tinglings which make his judgments on the position of the points as erratic at times as is the throw of a coin

In the figure, there is shown, in addition to the polygon based on experiment, a dotted theoretical polygon to which we now turn our attention. Since the probability of obtaining a head at one throw is $\frac{1}{2}$, the probability of obtaining seven heads in a throw of seven coins is $(\frac{1}{2})^7$ or $\frac{1}{128}$.

The probability that a certain coin will give a tail but all the others heads, is $(\frac{1}{2})^6 \times (1 - \frac{1}{2})$ which also equals $\frac{1}{128}$. As there are seven coins in all, each of which might give the only tail, the total chance of obtaining six heads and one tail is $\frac{7}{128}$. The probability that two specified coins out of the seven will give tails, the other five giving heads, is $(\frac{1}{2})^5 \times (1 - \frac{1}{2})^2 = \frac{1}{128}$. The number of ways in which the two can be

specified is 21. The total chance of obtaining five heads and two tails is, therefore, 21/128, obtained thus,

$$\frac{7!}{5!\,2!}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5=\frac{21}{128}.$$

It will be seen that the above probabilities of obtaining seven, six or five heads are the first three terms in the expansion of $(\frac{1}{2}+\frac{1}{2})^7$ The next term will similarly be found to be the probability of obtaining four heads, and so on. We thus obtain the following results.

Probability of 7 heads	1 in 128
" " 6 "	7 " "
" " 5 "	21 " "
" " 4 "	35 " "
" " 3 "	35 " "
" " 2 "	21 " "
" " 1 head	7 " "
" " no heads	1 " "
	128

It is these numbers which are shown by the dotted line in the figure.

In general, if p is the probability of an event succeeding, and q of it not succeeding, the respective chances of it succeeding

$$k,\ k-1,\ k-2,\ \dots\ 3,\ 2,\ 1\ \text{or}\ 0$$

times in k trials are given by the terms of the expansion of $(p+q)^k$.

In n groups of k trials therefore the most probable numbers of successes in the various groups are

$$n\left\{p^k+kp^{k-1}q+\frac{k(k-1)}{1.2}p^{k-2}q^2+\dots+q^k\right\}.$$

For example let the event be throwing either an ace or a six with a six-faced die, let six trials be made in each group, and let a thousand groups be tried. Then the above expansion is

$$1000\left\{\left(\frac{1}{3}\right)^6+6\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)+\frac{6.5}{1.2}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2+\frac{6.5.4}{1.2.3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3\right.$$
$$\left.+\frac{6.5}{1.2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4+6\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^5+\left(\frac{2}{3}\right)^6\right\}.$$

Therefore the number of times a throw consisting of *only* "aces" or "sixes" will occur in 1000 trials with six dice each time is most probably $\frac{1000}{3^6}=\frac{1000}{729}$ or once, to take the nearest integer. In the same way we get this table:

1000 throws of six dice each time

All 6 are "Aces or Sixes" on	$\frac{1000}{729}$ or approx	1 occasion	
Only 5 " " " " " "	$\frac{12000}{729}$ " "	16 occasions	
" 4 " " " " " "	$\frac{60000}{729}$ " "	82 "	
" 3 " " " " " "	$\frac{160000}{729}$ " "	220 "	
" 2 " " " " " "	$\frac{240000}{729}$ " "	329 "	(Mode)
" 1 is an ace or six "	$\frac{192000}{729}$ " "	264 "	
None are aces or sixes "	$\frac{64000}{729}$ " "	88 "	
		1000	

Here the *mode* is at two.

What is the *mean*? In calculating it we shall take the exact fractions in the above table, not the approximate integers. The mean is then obtained thus:

$6 \times 1000-729 = 6000-729$	
$5 \times 12000-729 = 60000-729$	$\text{Sum} = \frac{1458000}{729} = 2000$
$4 \times 60000-729 = 240000-729$	
$3 \times 160000-729 = 480000-729$	$\text{Mean} = \frac{2000}{1000} = 2$
$2 \times 240000-729 = 480000-729$	
$1 \times 192000-729 = 192000-729$	
$0 \times 64000-729 = \text{nil}$	

Here, therefore, the *mean* coincides with the *mode*.

In general we have from the binomial expansion

$$p^k + kp^{k-1}q + \frac{k(k-1)}{1.2}p^{k-2}q^2 + \dots,$$

Mean equals

$$kp^k + (k-1)kp^{k-1}q + (k-2)\frac{k(k-1)}{1.2}p^{k-2}q^2 + \dots$$

$$= kp\left\{p^{k-1} + (k-1)p^{k-2}q + \frac{(k-1)(k-2)}{1.2}p^{k-3}q^2 + \dots\right\}$$

$$= kp\,(p+q)^{k-1} = kp \text{ since } p + q = 1.$$

The *mode* can also be deduced from the expansion. The first term in the expansion, p, corresponds to k successes, the second term to $k-1$ successes, and so on, so that the term corresponding to any number m of successes is

$$\frac{k!}{m!\,(k-m)!}p^m q^{k-m}.$$

This term is made from the preceding term by multiplying the latter by

$$\frac{m+1}{k-m}\cdot\frac{q}{p}$$

and the succeeding term is made from it by multiplying by

$$\frac{m}{k-m+1}\cdot\frac{q}{p}.$$

It will therefore be the largest term if

$$\frac{m+1}{k-m}\,\frac{q}{p} > 1 > \frac{m}{k-m+1}\cdot\frac{q}{p},$$

whence

$$qm+q > pk-pm, \quad pk-pm+p > qm,$$
$$qm+pm > kp-q, \quad kp+p > qm+pm,$$
$$m > kp-q, \quad kp+p > m,$$

so that

$$kp-q < m < kp+p.$$

The greatest term, therefore, is that which corresponds to a number of successes between $kp-q$ and $kp+p$. The range thus indicated is unity since $p+q=1$. In the binomial expansion therefore the mean and the mode agree to an integer.

Standard Deviation of the Binomial Expansion

For the binomial expansion the standard deviation is equal to $\sqrt{(kpq)}$. This can be proved as follows

In the expansion of $(p+q)^k$ the first term p^k represents the probability of the specified event succeeding k times. The mean number of times it succeeds is kp, so that the deviation is $k-kp$ and the first term in σ^2 is $p^k \times (k-kp)^2$. Similarly the second term $kp^{k-1}q$ of the binomial represents the probability of $(k-1)$ successes, and the deviation is therefore $(k-1)-kp$, so that the second term in σ^2 is

$$kp^{k-1}q\,(k-1-kp).$$

Remembering that $k-kp=kq$ we get the following expansion for σ^2

$$\begin{aligned}
\sigma^2 &= (kq)^2 p^k + (kq-1)^2 kp^{k-1}q + (kq-2)^2 \frac{k(k-1)}{1.2} p^{k-2}q^2 + \dots \\
&= k^2q^2(p+q)^k - 2k^2q^2(p+q)^{k-1} + \left\{kp^{k-1}q + 4\frac{k(k-1)}{1.2}p^{k-2}q^2 + \dots\right\} \\
&= k^2q^2 \qquad - 2k^2q^2 + kq\,\{p^{k-1} + 2(k-1)p^{k-2}q + \dots\} \\
&= \qquad - k^2q^2 + kq\,\{(p+q)^{k-1} + (k-1)q(p+q)^{k-2}\} \\
&= \qquad - k^2q^2 + kq\,\{1 + (k-1)q\} \\
&= \qquad - k^2q^2 + kq\,(1 + kq - q) \\
&= kq(1-q) = kpq, \qquad \sigma = \sqrt{(kpq)}.
\end{aligned}$$

(6) THE NORMAL CURVE OF ERROR

We have seen that a binomial expansion gives a cocked hat figure very like the actual diagrams obtained by experiment. In the imaginary example which we considered, seven Djinns added or subtracted each a half unit to or from the quantity $13\frac{1}{2}$ which we were trying to measure. We then obtained measurements extending from 10 to 17 units, but always at the exact units.

In practice if we were handling a not easily measurable quantity we might well find our measurements range over this distance, but unless they were constrained to do so by some peculiarity of the method, they would not always occur at exact units, but at any distance. This case is covered if we imagine the number of Djinns (the factors of accidental error) to be much increased but the influence of each one made less: thus there might be 7000 Djinns each adding or subtracting a mere fraction of a unit, or ultimately an infinite number of them, each adding or subtracting an infinitesimal amount, just as an infinite number of the tiniest errors (we may well imagine) account for the variations in our experimental readings.

What would the binomial expansion then become? That is, what form does the expansion of $(\frac{1}{2}+\frac{1}{2})^k$ take when k becomes infinite and the terms are not unit distance apart but only an infinitesimal distance dx? The whole range covered by the expansion is kdx, and the term at either end which occurs when all the errors are either positive or negative, is at a distance $k\dfrac{dx}{2}$, from the middle, so that each elementary error is now $dx/2$ just as formerly it was half a unit.

The term which corresponds to l errors being positive and $k-l$ negative is

$$P = (\tfrac{1}{2})^l\,(\tfrac{1}{2})^{k-l}\,k!/\{l!\,(k-l)!\} \qquad \text{.....(1)},$$

and the net error, the abscissa of this point, is

$$x = \{l-(k-l)\}\,dx/2 \qquad \text{......(2)}$$

The next point, distant dx, is that corresponding to $l+1$ positive and $k-l-1$ negative errors, giving an abscissa of

$$\{l+1-(k-l-1)\}\,dx/2.$$

Its probability is

$$P+dP = (\tfrac{1}{2})^k\,k!/\{(l+1)!\,(k-l-1)!\}.$$

The ratio of these is

$$\frac{P+dP}{P} = \frac{k-l}{l+1} = 1+\frac{dP}{P}, \quad \therefore \frac{dP}{P} = \frac{k-2l-1}{l+1}.$$

But $l = \frac{x}{dx} + \frac{k}{2}$, from eqn. (2) for x above, so that

$$\frac{dP}{P} = -\frac{2(2x + dx)}{2x + (k+2)\,dx}.$$

Now we are going to make k, the number of atomic errors, equal to infinity, and we can therefore neglect the 2 in $k + 2$ and write

$$\frac{dP}{P} = -\frac{2(2x + dx)}{2x + k\,dx} = -\frac{4x}{2x + k\,dx} - \frac{2dx}{2x + k\,dx} \quad \ldots\ldots (3).$$

Therefore $\frac{dP}{dx}$, which is the quantity we require before we can integrate and obtain the equation to the continuous curve to replace the binomial cocked hat, is given by

$$-\frac{1}{P}\frac{dP}{dx} = \frac{4x}{2x\,dx + k\,dx^2} + \frac{2}{2x + k\,dx} \quad \ldots\ldots (4).$$

We must now consider the quantity kdx which gives the entire extent of scatter, the whole range. The number of errors k is infinite, we have assumed, and dx is infinitesimal The range kdx may then either be finite or infinite If we assume it to be finite, then kdx^2 will be infinitesimal and the above equation becomes

$$-\frac{1}{P}\frac{dP}{dx} = \frac{4x}{2x\,dx} + \frac{2}{2x + k\,dx} = \text{infinity},$$

when dx becomes infinitesimal. Or $\frac{dP}{dx} = -$ infinity, except when $P = 0$.

In this case the probability falls off infinitely quickly from the mean, and this is therefore a case of no scatter at all. If therefore we postulate an infinite number of elementary errors, we must allow a possible range of scatter of infinite extent, the only alternative being no scatter at all. But although the possible range is infinite, it will be found that at infinity the probability is infinitesimal, that is the larger errors do not in practice occur.

We take therefore

$$k\,dx = \infty,$$

$$k\,(dx)^2 = \text{a finite quantity*},$$

which we shall write $= 4\sigma^2$.

(σ will turn out presently to be our previous acquaintance the standard deviation.) We then have from eqn. (4)

$$-\frac{1}{P}\frac{dP}{dx} = \frac{4x}{2x\,dx + 4\sigma^2} = \frac{x}{\sigma^2} \quad \ldots\ldots (5),$$

* Here again the alternative ought to be investigated.

and on integrating,

$$\log P = -\frac{x^2}{2\sigma^2} + \text{constant},$$

or

$$P = Ce^{-x^2/(2\sigma^2)} \qquad \ldots\ldots(6).$$

This equation gives the probability P of any value of the error x occurring, and not only of any integral value, as the binomial does. The value of the constant C will be found presently, and it will also be shown that σ is, as has been already asserted, the same standard deviation which we already know in another guise. The general shape of such Normal Curves is shown by the example in Fig. 6, p. 43

Some properties of the Normal Curve

In the curve $P = Ce^{-x^2/(2\sigma^2)}$, P is the probability of the occurrence of an error x. Let us consider of what order of magnitude the quantities P are. In the first place we remember that the curve was found as the limit of the expansion $(\frac{1}{2} + \frac{1}{2})^k$, when k was made infinite. This curve however flattens out more and more as k is increased.

For example

$$(\tfrac{1}{2} + \tfrac{1}{2})^2 = \tfrac{1}{4} + \tfrac{1}{2} + \tfrac{1}{4},$$
$$(\tfrac{1}{2} + \tfrac{1}{2})^3 = \tfrac{1}{8} + \tfrac{3}{8} + \tfrac{3}{8} + \tfrac{1}{8},$$
$$(\tfrac{1}{2} + \tfrac{1}{2})^4 = \tfrac{1}{16} + \tfrac{4}{16} + \tfrac{6}{16} + \tfrac{4}{16} + \tfrac{1}{16}.$$

So that clearly when k becomes infinite, the ordinates of the curve, as we have so far considered it, flatten out so that they are all infinitesimal In other words the probability P of any exact value of x occurring is really infinitesimal though it varies from one x to another x. C must, therefore, be infinitesimal. Let us try putting

$$C = C'dx,$$

so that if C is of the same order of smallness as dx, the new quantity C' will be a *finite* constant. We have then

$$P = C'e^{-x^2/(2\sigma^2)}dx \qquad \ldots\ (7).$$

This means, in geometrical language, that instead of using a curve whose *ordinate* P measures the probability of the occurrence of x, we had better use a curve where this probability is measured by the *area* of an elemental rectangle contained by two ordinates enclosing x, and distant dx from one another (cf. Fig 4, p. 17). Such a curve will be similar in shape to the former but of *finite* ordinates.

We have assumed in the above argument that C' as there defined

is a finite quantity, that is that C is of the same degree of smallness as dx. That this is so may be shown by proceeding next to find the actual value of C' in the following manner.

It is clear that the sum of the probabilities of all errors must equal unity, for at any one measurement it is certain that some error or other must occur, if we include zero error in the mathematical sense. That is

$$\sum_{-\infty}^{\infty} P = 1.$$

Substitute for P from the equation (7) above and replace the sign of summation by integration. We thus obtain

$$C' \int_{-\infty}^{\infty} e^{-x^2/(2\sigma^2)}\, dx = 1 \qquad \text{......(8).}$$

Write $$x^2/(2\sigma^2) = z^2, \quad x\,dx = 2\sigma^2 z\,dz;$$

then the integral becomes

$$C' \int_{-\infty}^{\infty} e^{-z^2} \frac{2\sigma^2 z\,dz}{z\sigma\sqrt{2}} = C'\sigma\sqrt{2}.2B^* = C'\sigma\sqrt{(2\pi)} \quad \text{...} \quad \text{(9).}$$

Whence, since this equals unity, we have

$$C' = 1/(\sigma\sqrt{2}\sqrt{\pi}), \text{ a finite quantity} \qquad \text{..... (10).}$$

So that we can now write

$$P = \frac{dx}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)} \qquad \text{......(11).}$$

The ordinate of the curve defined on p. 35 is $y = P/dx$ so that its equation is

$$\mathbf{y = \frac{1}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)}} \qquad \text{.... .(12).}$$

Such a curve is called a probability curve. The probability of the occurrence of the value x is given by $y_x dx$ and the probability of x falling between a and b is given by $\int_b^a y\,dx$. The total area $\int_{-\infty}^{\infty} y\,dx$ equals unity.

If N measurements were made, and the errors were distributed according to such a curve, then the most probable number of times a deviation x occurred would be NP_x. A curve

$$y = \frac{N}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)}$$

is similar to a probability curve, but each ordinate is N times as tall†.

* The reference is to Integral B in Appendix II giving the values of a number of integrals of general use in the theory of probability

† Cf Keynes, *A Treatise on Probability*, London, 1921, p. 101 *et passim*. Keynes does not regard probability as identical with statistical frequency.

It is a *distribution* curve and its total area is not unity but N, and the integral from a to b represents not the probability of x falling between a and b, but the most probable number of times it would fall in this range out of N trials

Since this curve is symmetrical, the mean, mode and median coincide.

Let us find the standard deviation of such a set of measurements.

The sum of the squares of the deviations is given by

$$\int_{-\infty}^{\infty} \frac{N\,dx}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)} \times x^2$$

$$(\text{write } x^2 = 2\sigma^2 z^2, \quad x\,dx = 2\sigma^2 z\,dz)$$

$$= \frac{N}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-z^2} \times 2\sigma^2 z^2 \times \frac{2\sigma^2 z\,dz}{z\sigma\sqrt{2}} = \frac{N\,2\sigma^2}{\sqrt{\pi}} \times 2C^* = N\sigma^2.$$

The *mean square deviation* is $N\sigma^2/N = \sigma^2$

The *root mean square deviation*, or standard deviation, is therefore σ, so that we were justified in using this letter on p 34 when we wrote

$$k\,(dx)^2 = 4\sigma^2.$$

The quantity σ can be shown to have another significance in the normal curve, namely it is the distance from the centre to the point of inflection or point where the curve changes from convex to concave. The proof of this statement is as follows. At a point of inflection of a curve, the value of d^2y/dx^2 is zero In our case

$$y = \frac{1}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)},$$

$$dy/dx = -\frac{1}{\sigma^3 . \sqrt{(2\pi)}}\, x e^{-x^2/(2\sigma^2)},$$

$$d^2y/dx^2 = -\frac{1}{\sigma^3 . \sqrt{(2\pi)}} \left\{ e^{-x^2/(2\sigma^2)} - \frac{x^2}{\sigma^2}\, e^{-x^2/(2\sigma^2)} \right\}.$$

If this is to equal zero then x^2 must equal σ^2.

Finally, if we suppose the area enclosed between a distribution curve and the axis of x to be spinning round the axis of y, then the moment of inertia is such that we can consider the whole weight concentrated equally at the two points of inflection. In other words, σ is the "radius of gyration." For the area (or weight) of each vertical elementary column of the curve is $y\,dx$ and its distance from the axis of gyration is x.

The moment of inertia is therefore

$$\int_{-\infty}^{\infty} \frac{N}{\sigma\sqrt{(2\pi)}}\, e^{-x^2/(2\sigma^2)}\, x^2\,dx.$$

* See Integral C, Appendix II, p 202.

Write $x^2 = 2\sigma^2 z^2$ and the integral becomes

$$\frac{4\sigma^2 N}{\sqrt{\pi}} \times C^* = \sigma^2 N.$$

Therefore the N can be considered as concentrated at a distance σ from the centre, or $\frac{N}{2}$ at each point of inflection.

The Centroid of a Vertical Slice of a Normal Curve.

If c be the abscissa of the centroid of a slice bounded by ordinates at a and b then by definition of a centroid or centre of gravity,

$$\begin{aligned} c \times \text{area of slice} &= \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-x^2/(2\sigma^2)}\, x\, dx \\ &= \sigma^2 \left[\frac{e^{-x^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}\right]_b^a \\ &= \sigma^2 (y_a - y_b) \end{aligned}$$

For unit area and unit σ therefore the abscissa of the centroid of a slice is the difference of the bounding ordinates divided by the area of the slice (cf p. 128).

The Relation between Mean Variation and Standard Deviation, in the case of Normal Distribution.

The mean variation is the abscissa of the centroid of half the curve: and therefore by the above

$$\begin{aligned} \text{M.V.} &= \frac{\sigma^2 (y_0 - y_\infty)}{\frac{1}{2}} \\ &= 2\sigma^2 \left(\frac{1}{\sigma\sqrt{2\pi}} - 0\right) \\ &= \frac{\sqrt{2}}{\sqrt{\pi}}\, \sigma = 0{\cdot}8\sigma \text{ approx.} \end{aligned}$$

The Relation between Probable Error and Standard Deviation, in the case of Normal Distribution.

The probable error was defined on p. 22 as an arbitrary reduction of the standard deviation, viz.

$$\text{P.E.} = {\cdot}6745\sigma.$$

In the case of Normal Distribution a physical meaning can be given to

* See Integral C in Appendix II.

this quantity. Consider the number of cases which fall within the limits of the range $\pm\ 6745\sigma$. This number is

$$2 \times \frac{N}{\sigma\sqrt{(2\pi)}}\int_0^{6745\sigma} e^{-x^2/(2\sigma^2)}\,dx.$$

Values of the probability integral have been calculated and tabulated for all limits, and if such a table be consulted* it will be found that this quantity has the value $\frac{N}{2}$. That is to say, the probable error gives the range within which one-half the cases may be expected to fall It is more instructive to remember that half the cases may be expected to fall *outside* $\pm$ P.E

With skew distributions this meaning ceases to hold, and for these σ should be used. Indeed it is in general a better quantity to employ.

(7) ON FITTING A NORMAL CURVE TO DISTRIBUTION DATA

Consider the experiment of bisecting a line of which the data are given on p 15. The histogram of Fig. 3, p. 16 represents these data, and shows the density with which these points occur in each part of the range. We wish to replace this stepwise figure by a smooth Normal Curve which will give us at each point of the range the theoretical proportionate density of the bisection points at that spot, and we want this Normal Curve to be the most probable which can be based on the given data.

The equation of the required curve is

$$y = \frac{N}{\sigma\sqrt{(2\pi)}}\, e^{-(x-a)^2/(2\sigma^2)}$$

where N is the number of experiments made, x is measured in mms, and $(x - a)$ is a quantity measured from some central point of the data In the theoretical curve (which is symmetrical) a is the point where mean, mode and median coincide. In the data however the mean is 60·13 and the median 60 while the mode is unknown. What value shall we give to a so as to obtain the best fitting curve?

To answer these questions it is necessary to consider what we mean by *best fitting curve*.

The meaning of "best fitting curve" is really the same as the meaning attached to the phrase "most probable theory" By the best theory of any set of data we mean the theory from which the observed data could have chanced to spring with a greater probability than would

* E g. the first table in Pearson's *Tables for Statisticians and Biometricians*

be the case with any other theory. Take a simple example. Suppose a bag is known to contain a large number of equal-sized balls and nothing else: but the colours of the balls are entirely unknown. Experiments have been carried out to learn something about their colours, each experiment taking the form of extracting one ball, noting its colour, and replacing it. Ten such experiments have been made and the results are as follows:

5 black balls,
3 white balls,
1 red ball,
1 green ball.

What is the best theory of the composition of the bag?

The best theory on the facts as given, is that the bag contains $\frac{5}{10}$ths black balls, $\frac{3}{10}$ths white balls, $\frac{1}{10}$th red balls, and $\frac{1}{10}$th green balls.

Suppose however that another theory was advanced, namely that the bag contained $\frac{4}{12}$ths black balls, $\frac{4}{12}$ths white balls, and $\frac{1}{12}$th each of red, green, yellow and blue balls. How should these two theories be compared?

The proper plan is to find the probability, on each theory, that ten dips would result in what was actually observed. Then that theory is best for which this probability is greatest. Let us take first the theory which we assert to be the best. The probability, on that theory, of drawing five black, three white, one red and one green ball (the order being immaterial) is

$$\left(\frac{5}{10}\right)^5\left(\frac{3}{10}\right)^3\cdot\frac{1}{10}\cdot\frac{1}{10}\cdot\frac{10!}{5!\,3!\,1!\,1!}=0{\cdot}042525.$$

The corresponding probability on the other theory is

$$\left(\frac{4}{12}\right)^5\left(\frac{4}{12}\right)^3\cdot\frac{1}{12}\cdot\frac{1}{12}\cdot\frac{10!}{5!\,3!\,1!\,1!}=0{\cdot}005335.$$

The former theory is therefore the better of the two. If from other sources we knew however that the bag did contain balls of six colours, then the first theory would be ruled out, and the second theory would have to be compared with other six-colour theories. The reader might, for example, compare it with the theory that the bag contains $\frac{5}{12}$ths black, $\frac{2}{12}$ths white, $\frac{2}{12}$ths red and $\frac{1}{12}$th each of green, blue and yellow balls.

Turn now to our actual experiments on bisecting a line. The 29 experiments are like 29 dips into a bag, resulting in the numbers on p. 15 being drawn. From our general consideration of the problem we believe that the numbers in the bag form a continuum, that is we

think that another bisection mark might fall anywhere within the range, and not only at points already struck, though of course our measurements are only being made to the nearest tenth-millimetre. Secondly we think that a curve of the form

$$y = \frac{N}{\sigma\sqrt{(2\pi)}} e^{-x^2/(2\sigma^2)}$$

will express this continuum. The problem is, *which* curve of this form is the best fitting one. This is decided in exactly the same way as was the above example of the coloured balls For each curve find the probability that the actually observed distribution may have arisen from it by random sampling. Then that curve is the best for which this probability is greatest.

The actual process of trial and error giving now this now that value to σ and a would take too long, and we find the optimum values by the usual process of the differential calculus. Let us do so in a quite general manner, taking n values (instead of 29 as here).

Given a distribution following the law

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-a)^2/(2\sigma^2)},$$

what is the probability Q that the n special values

$$x_1, x_2, x_3, x_4, \dots x_\lambda, \dots x_n$$

will be obtained in a set of n trials (the order being immaterial)? The probability of x_λ is

$$\frac{dx}{\sigma\sqrt{(2\pi)}} e^{-(x_\lambda - a)^2/(2\sigma^2)}$$

and the required probability is equal to the continued product of a number of such expressions, in which λ takes the values 1 to n successively, multiplied by $n!$ because the order is immaterial.

We therefore have

$$Q = n!\left(\frac{dx}{\sigma\sqrt{2\pi}}\right)^n e^{-S(x_\lambda - a)^2/(2\sigma^2)}.$$

This probability we wish to make a maximum by choosing the best values of σ and a. We must therefore put

$$dQ/da = 0$$

and

$$dQ/d\sigma = 0.$$

Now $$\frac{dQ}{da} = n!\left(\frac{dx}{\sigma\sqrt{2\pi}}\right)^n e^{-S(x_\lambda - a)^2/(2\sigma^2)} \times \frac{S(x_\lambda - a)}{\sigma^2},$$

and since this equals zero we must have

$$S(x_\lambda - a) = 0,$$
$$\therefore S(x_\lambda) = S(a) = na,$$
$$a = S(x_\lambda)/n.$$

That is, a must be the mean of the observations Turning now to the second equation we have (omitting from the first the factor $dx/\sqrt{(2\pi)}$, which is independent of σ)

$$\frac{d}{d\sigma}\left\{\frac{1}{\sigma^n} e^{-S(x_\lambda - a)^2/(2\sigma^2)}\right\} = \frac{e^{-S(x_\lambda - a)^2/2\sigma^2}}{\sigma^{n+3}}\left\{S(x_\lambda - a)^2 - n\sigma^2\right\} = 0.$$

Therefore $$n\sigma^2 = S(x_\lambda - a)^2,$$
$$\sigma^2 = S(x_\lambda - a)^2/n,$$

i.e. σ is the standard deviation of the readings.

We find therefore that to obtain the best fitting curve we must find a the value of the mean of the observations, and σ the value of the standard deviation. In the case in question

$$a = 60{\cdot}13,$$
$$\sigma = 1{\cdot}38,$$

so that $$y = \frac{29}{1{\cdot}38\sqrt{(2\pi)}} e^{-(x-60\ 13)^2/(2\times 1\ 38^2)}$$

where x is in mms. This curve is drawn on the adjoining figure. Calculations, such as those required to find a number of ordinates of the above curve for the purpose of drawing it, are best performed in tabular fashion. For example, the present calculation might be arranged as follows, and the reader should calculate one or two ordinates for practice by this means

(a)	(b)	(c)	(d)	(e)	(f)
x in mms.	$x_1 = x - 60\ 13$	$\frac{x_1^2}{2\times 1\ 38^2}$	Column c × log e	Reciprocal of Antilog column d	Column $e \times \frac{29}{1\ 38\sqrt{(2\pi)}} = y$

This arrangement is suitable for an approximate calculation using a ten inch slide rule, from which the logarithms are also taken. The

accuracy thus attained is quite as great as the extent of the experiment justifies. If logarithm tables, or calculating machines, were to be used, somewhat different tabular arrangements would be required.

Some of the calculation in the above table has, however, once and for all been done and printed in tables of the probability curve The best of these for our purpose is Sheppard's table, printed as Table II in Professor Pearson's *Tables for Statisticians and Biometricians.*

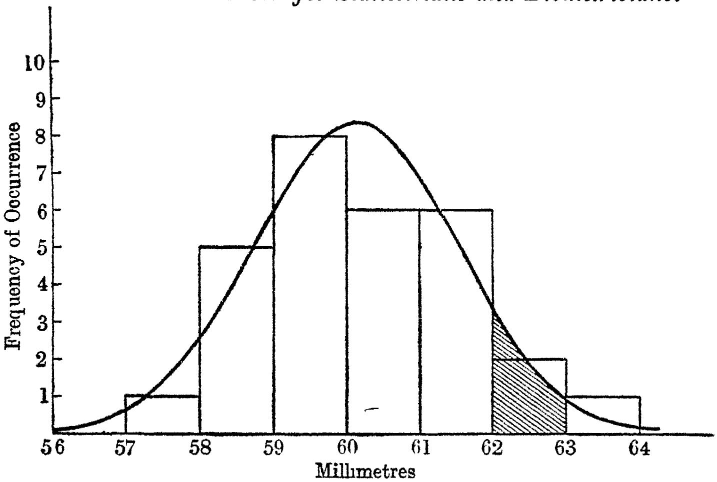


Fig. 6 A normal curve fitted to the bisection data. The histogram is also shown

*Fitting a Normal Curve to the Bisection Data with the Aid of Sheppard's Tables**

I	II	III	IV	V
x in mms	$x' = x - 60\ 13$	Sheppard's x $x'/1\ 38$	Sheppard's z	$y = 29z/1\ 38$
60 13	0	0	·3989432	8 38
60	0 13	0 09	397	8 34
59 5	0·63	0 46	359	7 54
59	1 13	0 82	285	5 99
58 5	1 63	1 18	199	4 18
58	2 13	1 54	122	2 56
57	3 13	2 27	030	0 63
56	4 13	2 99	005	0 11

The other half of the curve can be drawn by symmetry No interpolation is here used in the tables. A slide rule, or Crelle's Calculating Tables, can be used for the multiplications

* Of course an experiment of only 29 observations does not justify any curve fitting at all, as the accuracy is not sufficient, the sample not being large enough But a *short* example is necessary for explanatory purposes in a text book.

We are at present only concerned with Sheppard's first and fifth columns headed x and z respectively. These are connected by the relationship

$$z = \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2},$$

that is they give a curve with $N = 1$ and $\sigma = 1$. The x of Sheppard's table is, therefore, obtained from our x by division by σ, and his z needs to be multiplied by N/σ. Our calculation then takes the form shown above Fig 6.

It must be clearly understood that there are two parts in any curve fitting problem Firstly there is the decision as to what kind of curve is to be used, and secondly the finding of the best fitting curve *of this kind*. If it were merely a question of getting a curve to fit the particular data of one problem, then a curve could always be drawn to fulfil the conditions *exactly*. It is only because general considerations dictate that the curve shall be of a certain kind, that an exact fit cannot be obtained and the best fit has to be found.

(8) THE METHOD OF LEAST SQUARES

The principles adopted in the last section are those which underlie the Method of Least Squares, which is employed to find the best solutions of a set of linear equations which are more numerous than the unknowns, and slightly inconsistent with one another. To illustrate the principle take three equations for two unknowns,

$$\begin{aligned} ax + by + c &= 0, \\ a'x + b'y + c' &= 0, \\ a''x + b''y + c'' &= 0. \end{aligned}$$

The quantities a, a', a'', b, etc. having been measured, the equations are found not to come to zero for any values of x and y, but to give "residuals" v, v', and v'' thus·

$$\begin{aligned} ax + by + c &= v, \\ a'x + b'y + c' &= v', \\ a''x + b''y + c'' &= v''. \end{aligned}$$

Now the presence of these residuals v may be assumed to be due to numerous small errors in the coefficients a, b, c, and the distribution of any v, were one of these equation-observations to be repeated many times, will, it may be assumed, be a Normal Curve. The probability of occurrence of v will therefore contain as chief factor e^{-v^2} and the

probability of occurrence of v, v' and v'' (presuming them independent) will contain the factor

$$e^{-v^2} \times e^{-v'^2} \times e^{-v''^2}, \quad \text{or} \quad e^{-S(v^2)},$$

where $$S(v^2) = v^2 + v'^2 + v''^2.$$

To make this probability a maximum we must make $S(v^2)$ a minimum, whence the name Least Squares. The conditions for $S(v^2)$ a minimum are

$$dS/dx = dS/dy = 0,$$

and also (though in practice it is not necessary to find them) the second differentials must be negative. We get at once

$$\frac{d}{dx}\left\{(ax + by + c)^2 + (a'x + b'y + c')^2 + (a''x + b''y + c'')^2\right\}$$
$$= 2(a^2 + a'^2 + a''^2)x + 2(ab + a'b' + a''b'')y + 2(ac + a'c' + a''c'') = 0,$$

and similarly for y.

The two equations thus reached are called Normal Equations. There will of course always be as many of them as there are unknowns If the original equations were not of equal reliability or weight they must of course first be multiplied by their weights. The rule of Least Squares is therefore: *To obtain the Normal Equation for any unknown x, multiply each equation by its weight w, and by the coefficient of x in that equation, and then add all the equations together*. The Normal equations in our example are thus.

$$S(a^2w)\,x + S(abw)\,y + S(acw) = 0,$$
$$S(abw)\,x + S(b^2w)\,y + S(bcw) = 0,$$

whence unique values of x and y can be found.

Note, 1924 The work of Keynes on probability had not been published when this chapter was written, or a number of points might have been put differently For instance Keynes points out that odds at betting depend not only on the chance of success but also on the market available. And in his ch. VIII, on his pages 94 and 100, and elsewhere, are many fundamental questions which, though they could not be discussed, ought perhaps to have been mentioned here.

CHAPTER III

THE PSYCHOPHYSICAL METHODS

Experimental methods and mathematical processes—The method of limits—The method of average error—The constant method—Difference thresholds and the probability of a judgment of a certain category

(1) EXPERIMENTAL METHODS AND MATHEMATICAL PROCESSES

THE experimental determination of absolute and difference thresholds or limina is complicated and difficult. A considerable number of physical and psychological, and, it may be added, mathematical, factors is involved, of varying relative importance in different cases. The result is that different methods of procedure have been found most suitable for different cases. These methods have been traditionally grouped under three (or four) distinct headings, and called the Psychophysical Methods. They are

(1) the Method of Limits (Method of Minimal Changes),
(2) the Method of Average Error (Method of Production),
(3) the Constant Method (Method of Right and Wrong Cases).

A fourth method is generally added to the list, viz.:

(4) the Method of Equal Appearing Intervals, or Method of Mean Gradations, but this is really no new method. It owes its special name to the nature of the task which it fulfils, viz. the determination of equal-appearing (*uebermerklich*) sense-distances as distinguished from just perceptible (*ebenmerklich*) sense-distances. The method which it employs falls under one or other of the first three headings.

There are two things, essentially different from each other, which are commonly confused under this one heading "psychophysical methods," namely the *methods* of experimenting in order to obtain data, and the *processes* of calculation after the data have been collected. To avoid this confusion the words "method" and "process" will be employed throughout this book in the way indicated by their use in the above sentence. It is urged that their general adoption would be advantageous. The cause of the confusion is to be found in the historical development of the subject, for with each of the methods of experi-

menting a process of calculation was associated and the one name was given to both.

The experimental methods of determining thresholds may be divided into two main groups:

1. Methods in which the stimulus is altered continuously until in the opinion of the subject it fulfils some given condition.
2. Methods in which various values of the stimulus are separately submitted by the experimenter to the subject who expresses a judgment on each of them, classifying them into two or more categories.

The methods belonging to the second group may in turn be classified according to the order in which the stimuli are submitted to the subject. The second group is thus subdivided as follows.

2 (*a*). Methods in which the order of succession of the stimuli is irregular, non-consecutive

2 (*b*) Methods in which the order of succession of the stimuli is consecutively (i) ascending or (ii) descending.

Under 2 (*a*) comes the Method of Right and Wrong Cases. Under 2 (*b*) come the Method of Minimal Changes and the Method of Serial Groups*. The latter is a special case of the Method of Minimal Changes, in which each value of the stimulus is submitted a number of times before the next consecutive value is submitted. A corresponding method under 2 (*a*) is the Method of Non-Consecutive Groups, in which groups are taken in an irregular order. It is customary in the Method of Serial Groups to introduce among the stimuli an equal number of "catch" cases, in which no stimulus (or stimulus difference) at all is given. This might also be done in other methods. Each method can be further subdivided according as the subject is or is not warned beforehand of the kind of succession to expect.

The differences which result from the use of these various "methods" are clearly due to their different psychological effects upon the subject. But we have also to take into consideration the differences in the "processes" of calculation. These, of course, are purely mathematical and ignore the subject altogether.

* We adopt here the convenient name suggested by G M Stratton who first described the method (*Psychol Rev* 1902, IX. pp 444—447). It had, however, been independently discovered and employed by W. McDougall four years previously during his stay in Murray Island (*Rep. Cambridge Anthrop. Expedition to Torres Straits*, Cambridge, 1903, II. pp. 190—193).

(2) THE METHOD OF LIMITS

In using this method for the determination of difference limina the following mode of procedure is adopted. The variable stimulus V is first made equal to or slightly larger than the standard S, and then increased step by step by small increments until the subject finds it just perceptibly greater than S. V is increased still more, and then gradually diminished until it just ceases to appear greater than S. The mean of the two values of $V—S$ thus obtained is the upper difference limen, T_u.

Four values, instead of two, might also be obtained, viz. for (i) V just not perceptibly greater than S, (ii) V just perceptibly greater than S, (iii) V just perceptibly greater than S, (iv) V just not perceptibly greater than S; (i) and (ii) being for ascending values of V, (iii) and (iv) for descending. T_u will be the mean of these four values of $V—S$.

The lower difference limen, T_l, is obtained in a similar way. Both limina may be obtained in the same series of experiments by the "method of complete descents and ascents."

A series of determinations of each limen is made and the average taken. A measure of scatter (mean variation, say, or standard deviation) is also calculated. Absolute thresholds may also be found by this method.

The number and size of the increments employed must be adjusted to the particular conditions of the experiment. The subject of the experiment should be given a certain amount of preliminary practice before being started upon the work, and introspective reports should be asked of him.

Among the various possible sources of error which deflect the subject's judgment both in this and in the other psychophysical methods, two are of special importance. They arise from the temporal and spatial arrangement of the compared stimuli, and are called the "time error" and "space error" respectively. Thus, in a determination of the difference limen for sound-intensities, the two stimuli, S and V, cannot be presented simultaneously. One must precede the other, and in this way it may produce a slight degree of fatigue which causes an over-estimation of the other, or it may produce the reverse effect of sharpening the attention to the second. Thus a time error arises. Again, in experiments with brightness-intensities or visual extents, where the stimuli can be presented simultaneously, a space error arises from the fact that the one stimulus must be presented either to the right or to the left of the other stimulus, and the subject's judgment varies accordingly. In

experiments with lifted weights both sources of error may be involved. These so-called *constant errors* may be approximately neutralised by arranging that in the course of the experiment the standard shall precede the variable or stand to the right of the variable in half the cases and follow or stand to the left in the other half—the time or space order, or both, being of course changed quite at random in the successive limen determinations. A better plan is to evaluate the limen, and its scatter, for each time and space order separately. This gives us the values of the time error and the space error, which are of interest for their own sakes. Fechner's theory of these errors and their measurement is based upon the assumption that the time and space orders of the two stimuli, S and V, exert an influence upon the result which is equivalent to a definite increase or diminution in the stimulus-value of one or the other, and thus increase or diminish the value of $S \sim V$ by the amount of the time error e_1, or that of the space error e_2, or by the sum or the difference of these two. The time error is positive or negative according as the effect of the time order is to increase or diminish the apparent value of the *first-presented* stimulus. The space error is positive or negative according as the effect of the space order is to increase or diminish the apparent value of the *left-hand* stimulus.

Four principal cases of time and space order are possible, and are conventionally numbered as follows*.

I	Standard	presented	first	and to the	right,
II	,,	,,	second	,,	,,
III	,,	,,	first	,,	left,
IV	,,	,,	second	,,	,,

Employing these numbers as suffixes, we have the equations

$$\begin{aligned} T_{u_{\mathrm{I}}} &= T_u - e_1 + e_2, & T_{l_{\mathrm{I}}} &= T_l + e_1 - e_2, \\ T_{u_{\mathrm{II}}} &= T_u + e_1 + e_2, & T_{l_{\mathrm{II}}} &= T_l - e_1 - e_2, \\ T_{u_{\mathrm{III}}} &= T_u - e_1 - e_2, & T_{l_{\mathrm{III}}} &= T_l + e_1 + e_2, \\ T_{u_{\mathrm{IV}}} &= T_u + e_1 - e_2, & T_{l_{\mathrm{IV}}} &= T_l - e_1 + e_2, \end{aligned}$$

whence

$$T_u = \frac{T_{u_{\mathrm{I}}} + T_{u_{\mathrm{IV}}}}{2} = \frac{T_{u_{\mathrm{II}}} + T_{u_{\mathrm{III}}}}{2} = \frac{T_{u_{\mathrm{I}}} + T_{u_{\mathrm{II}}} + T_{u_{\mathrm{III}}} + T_{u_{\mathrm{IV}}}}{4},$$

and a similar expression holds for T_l; and

$$\begin{aligned} 2e_1 &= T_{u_{\mathrm{II}}} - T_{u_{\mathrm{I}}} = T_{u_{\mathrm{IV}}} - T_{u_{\mathrm{III}}} = T_{l_{\mathrm{I}}} - T_{l_{\mathrm{II}}} = T_{l_{\mathrm{III}}} - T_{l_{\mathrm{IV}}}, \\ 2e_2 &= T_{u_{\mathrm{I}}} - T_{u_{\mathrm{III}}} = T_{u_{\mathrm{II}}} - T_{u_{\mathrm{IV}}} = T_{l_{\mathrm{III}}} - T_{l_{\mathrm{I}}} = T_{l_{\mathrm{IV}}} - T_{l_{\mathrm{II}}}. \end{aligned}$$

* G. E. Müller, *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, 1904, pp. 67, 71.

The errors due to expectation, habituation, fatigue, etc, are neutralised or at least reduced to a minimum by determining the limen by means of both ascending and descending values of V and averaging, and by other special precautions in applying the method.

The Mathematics of the Method of Limits

It is to Professor F. M. Urban that we owe the most complete discussion of the mathematical foundations of this method*.

Let the variable stimulus V have the values

$$s_1, s_2, s_3, \ldots s_n.$$

Then for constant experimental conditions there will be, Prof Urban assumes, for each stimulus a probability that a certain judgment (say *Greater*) will be given Let these probabilities be

$$p_1, p_2, p_3, \ldots p_n.$$

If the stimuli were arranged in order of magnitude beginning with the smallest, then these probabilities will also be so arranged. Let us further write q for the probability that this judgment will *not* be given. Then for each q and p we have

$$1 - p = q.$$

Consider ascents first. A stimulus s is noted as a just perceptible point if the answer *greater* is given at s and was not given at any lower point in that series. This is a compound event, and its probability is the product of a number of q's for the lower stimuli where *greater* was not the answer, and one p for the stimulus s itself where the answer *greater* is returned. Let P represent the probability that s will be noted as one reading of the just perceptible point, then we have the set of equations

$$\begin{aligned} P_1 &= p_1, \\ P_2 &= q_1 p_2, \\ P_3 &= q_1 q_2 p_3, \\ \cdots\cdots\cdots & \;\cdots \\ P_n &= q_1 q_2 \cdots q_{n-1}\, p_n, \end{aligned}$$

and the mean of the just perceptible points will be

$$T = s_1 P_1 + s_2 P_2 + s_3 P_3 + \ldots + s_n P_n = S\,(sP).$$

The standard deviation will be given by

$$\sigma^2 = S\,\{(s - T)^2\, P\} = S\,(s^2 P) - T^2.$$

* See "Die psychophysischen Massmethoden," *Archiv f. d. ges. Psychologie*, 1909, XV. p 289; "On the Method of Just Perceptible Differences," *Psychol Rev.* 1907, XIV. p 244; *The Application of Statistical Methods to the Problems of Psychophysics*, Philadelphia, 1908.

An example will make the whole of this much clearer. Weights of 84, 88, 92, 96, 100, 104 and 108 grams were compared, by lifting, with a standard weight of 100 grams, and the replies given, which referred to the variable weight, were *heavier, equal* and *lighter*. They were recorded as follows.

84	88	92	96	100	104	108
e	l	h	h	e	h	h
l	h	e	h	h	h	h
l	l	l	h	h	h	h
l	l	l	e	e	h	h
l	l	e	h	h	h	h

and so on for 400 rows, the letters h, e and l being the initial letters of the answers given The five rows shown *in extenso* give five readings of the just perceptibly heavier point, viz 92, 88, 96, 104 and 96 grams The 400 of these obtained from the 400 rows were distributed as follows (Urban's Subject I)

Grams	84	88	92	96	100	104	108	
Frequency	0	7	30	76	106	169	12	400 in all.

The mean of these points is **100·36**, which is therefore the *directly observed* threshold of just perceptible positive difference (It should be noted that a time error is included which accounts for the nearness of this point to the standard) The standard deviation is 4 35 grams.

The frequency with which the answer *heavier* was given at each stimulus can also be found from the records The application of Urban's Formula, based on these frequencies p, is then as follows

Grams	s	p	q	q products	P	Ps	Ps^2
84	−4	·0022	9978	9978	0022	− 0088	0352
88	−3	0200	9800	9778	0200	− 0600	1800
92	−2	0889	9111	·8909	0869	− 1738	3476
96	−1	·2222	7778	6929	·1980	− 1980	1980
100	0	4133	5867	4065	2864		
104	1	8956	1044	0424	3641	3641	·3641
108	2	9400	0600	0025	0399	0798	1596
112	3				0025	0075	0225
			sums	4 0108	1 0000	4514	1 3070
			grams per working unit	4		− 4406	0001 = T^2
				16 0432		T = ·0108	1 3069 = σ^2
			Origin*	84	working unit	4	1 14 = σ
				100 0432		0432	4 working units
					origin	100	4 56 grams = σ
						100 0432	

* See explanation in text to follow

The *calculated* point of just perceptible positive difference is therefore **100·04** grams and the standard deviation 4·56 grams It will be seen that a working origin has been taken at 100 grams, and a working unit equal to 4 grams, to simplify the arithmetic. The column headed *q products* is the continued product of the q's from the 84 end. The P column is most quickly formed as the differences of successive numbers of the product column. The P's rise to a maximum and then sink again. They represent in fact the cocked hat distribution of the just perceptibly heavier points, and should be compared with the actual distribution of the latter The values P really give the distribution which would be found were an infinite number of ascents to be made, *the probabilities of the answer heavier at the different points remaining throughout constant at the actual frequencies which are found in these* 400 *ascents.*

It will be noticed that the P's do not add up to unity unless the amount 0 0025 is included. This quantity gives the probability that an ascent should be made without obtaining any answer *heavier*. In the example these "just perceptibly greater" points are centred at the weight 112, since it is necessary to make some assumption as to their position. Strangely enough, Professor Urban himself omits them in his calculations, causing in some cases quite an appreciable error. This difficulty about the "tail" of the P distribution would not have been necessary had the experiments been extended to a point where all the answers were *heavier*. The peculiar difficulty about this is however that the psychological conditions would thereby be changed*. The "tail" difficulty will therefore always be present in threshold measurements. It causes mathematical troubles in all the methods except only in the process of calculation used in the Constant Method, and will be frequently referred to in this and the succeeding chapters.

If the steps between the stimuli are equal, as here, a saving in arithmetic which escaped Professor Urban's notice can be effected by using the summation method described on page 19. The sum of the q products gives the distance of the threshold from the end stimulus. If the equal increments between the stimuli are not unity but x, the sum of the q products must first be multiplied by x The standard deviation can also be obtained by a further application of the summation process; the details are left to the reader who should consult the next chapter. The consideration of this point here might lead us too far from the main argument.

* See *inter alia* the article "On Judgments of Like," Frank Angell, *Am. Journ. Psychol.* 1907, XVIII. p. 253.

In the same experiment the points of *just not perceptible* difference were distributed as follows:

Grams	84	88	92	96	100	104	108	
Frequency	0	8	36	99	200	35	22	400 in all

Of these the mean is **98·84** grams, and the standard deviation 4·02 grams. The application of Urban's Formula in this case gives the following results:

Grams	s	p	p products	P	Ps	Ps^2
80	−5			0000	0000	0000
84	−4	0022	0000	0001	− 0004	0016
88	−3	0200	0001	0068	− 0204	0612
92	−2	0889	0069	0704	− 1408	2816
96	−1	2222	0773	2707	− 2707	·2707
100	0	4133	·3480	·4939		
104	1	8956	8419	0981	0981	·0981
108	2	9400	9400	·0600	1200	2400
			2 2142	1 0000	2181	9532
		working unit	4		− 4323	0459 = T^2
			8 8568	T =	− 2142	9073 = σ^2
		origin	108	working unit	4	·9525 = σ
			99 1432		− 8568	4 working unit
				origin	100	3 8100 grams = σ
					99 1432	

The mean threshold is therefore

by the direct method	(100 36 + 98 84)/2 = 99 60 grams,
by Urban's Formula	(100 04 + 99 14)/2 = 99 59 grams

Professor Urban's own calculated values are slightly different, partly because he neglects the tail of the distribution as above explained, partly because of an arithmetical error.

The practical applications of Urban's Formula are few, though it is occasionally useful in cases where the direct observation of the just perceptible points is impracticable*. Its great value lies in the conclusions which it enables us to draw. In the first place it shows clearly that the order in which the stimuli occur has no effect on the value of the threshold from a mathematical point of view, for the sum of the quantities sP is independent of their order, and the P's themselves are equally independent thereof. This does not of course mean that a change of order of the stimuli will have no psychological effect. There is no doubt that the effect of a non-consecutive series of stimuli will be very

* See e g. G. H Thomson, "Changes in the Spatial Threshold during a Sitting," *Brit. Journ. Psychol.* 1914, vi p 438.

different from that of an ascending or descending series, inasmuch as the latter will arouse feelings of expectation and the like Professor Urban therefore recommends that except when it is specially desired to study the results of ascents or descents the stimuli should be arranged in non-consecutive order In fact it will be seen that the above described simple mathematical process can also be applied to data collected by the Method of Right and Wrong Cases We have here an example of the distinction between methods of collecting data on the one hand and processes of calculation on the other The Limiting Process of calculation can be applied to the Constant Method of collecting data.

Professor Urban however, although recommending non-consecutive stimuli, does not recommend that the data should be collected in the way usual in the Constant Method if it is intended from the outset to apply the Limiting Process On the contrary, instead of keeping the stimuli constant throughout the experiment, he advises a frequent change of stimulus-series. For he found, from consideration of the results of interpolating and of irregularly spacing stimuli, that the final values obtained by the Limiting Process are dependent upon the particular stimuli used, in a mathematical way. To free results from this bias as far as possible, he recommends the following procedure The steps in any one series should be of equal size, and the starting point of the steps should be altered from series to series until the whole area has been covered. For example, if the spatial threshold were being investigated, the stimuli used might be*

First series	0, 1, 2, 3, 4, 5, 6, 7, 8 centimetres;
second series	0 1, 1 1, 2 1, 3 1, 4 1, 5 1, 6 1, 7 1, 8 1;
third „	0 2, 1 2, 2 2, 3 2, 4 2, 5 2, 6 2, 7 2, 8 2;
.. ..	. . ..
tenth „	0 9, 1 9, 2 9, 3 9, 4 9, 5 9, 6 9, 7 9, 8 9

The series might of course follow each other in any order. When this is done, Professor Urban shows that the final limen arrived at is the point where the probability of the response in question is one-half. That is, the Limiting Method finds the *median* threshold†

A modification of the Method of Limits, which effects a more complete elimination of the expectation error by the use of "catch experiments," where the variable given is equal to the standard, and which

* See G H Thomson, "Changes in the Spatial Threshold during a Sitting," *Brit. Journ. Psychol.* 1914, VI pp 435—6, where this plan is followed.

† For other discussions of the mathematics of the Limiting Process and allied subjects the reader may consult articles by Wirth, Urban, Lipps in the *Archiv f. d. ges Psychologie*, etc.

also collects data at a quicker rate than the ordinary Method of Limits, is that known as the *Method of Serial Groups**.

Each fixed value of the variable stimulus is presented with the standard ten times, not in immediate succession, but interspersed at random among ten other values of V equal to S. The percentage of correct answers given by the subject is noted, and the experimenter passes on to the next value of V which is presented along with catch stimuli in a similar manner. The value of V which, as presented in this way, gives 80 % right answers, is arbitrarily chosen as measuring the limen. Of course this choice of 80 % makes the method measure a totally different point from the 50 % point measured by the Method of Limits, but this is not essential to the method as a method of collecting data† As such it is a very convenient one to use in measuring a large number of subjects in "mental test" experiments, or with primitive people, where economisation of time is essential

The mathematical theory of this method is similar to that of the Method of Limits‡ It has been shown that *mathematically* the Method of Limits is superior, and that the mathematical disadvantages increase with the size of group taken For this reason groups of four have been suggested instead of ten§. Further, the arbitrary use of the 80 % point obscures comparison with the results of other methods The 50 % point would be better although it does not allow so large a number of subjects to be measured in a short time as does the 80 % point, if, as is assumed, each descent is stopped as soon as the required point is reached, and a new descent begun. With groups of four, the 75 %, 50 % and 25 % points can all be noted if time permits of complete descents and ascents (thus giving both the median limen and a measure of scatter, the interquartile range), while if time pressed, the 75 % point would be sufficiently comparable with the 80 % point of previous experiments.

As in the Method of Limits, the mathematical foundations are unaltered if the groups cease to be in serial order We thus arrive at

* See *Text Book of Experimental Psychology* C S Myers, Cambridge, 1911, p 196

† See G M. Stratton (*Psychol. Review*, 1902, IX pp. 444—447), who says "a detail like this, as well as the exact number of experiments that may best form a group, might well be considered as subject to revision in the light of further experience and not as an essential part of the method"

‡ G. H. Thomson, "A Comparison of Psychophysical Methods," *Brit Journ. Psychol.* 1912, V. p. 212

§ "An Inquiry into the Best Form of the Method of Serial Groups," *Brit Journ. Psychol.* 1913, V. pp. 398—416, and "The Probable Error of Urban's Formula," *ibid.*, 1913, VI. pp. 217—222.

a *Method of Non-Consecutive Groups**, which has been held by one experimenter to be the best method of collecting data

It may be pointed out in passing that the Method of Groups is one which is naturally employed in other branches of mental measurement as well as in psychophysics For example, the widely known Binet Tests are given in groups beginning at a group designed to suit a child of very young age, and are proceeded with until a certain percentage of passes is obtained at some group. Usually however modifications in the form of marks for tests passed above the critical group are admitted. There is no doubt but that the mathematical foundations of many of these devices require examination in the light of the theory of probability†

It may also be pointed out that as methods of collecting data the Method of Serial Groups, and that of Non-Consecutive Groups, especially the latter, approach the principle of the constant method yet to be described, and indeed their data can very well be handled by the processes in use in the latter method.

(3) THE METHOD OF AVERAGE ERROR

In this method the subject is required to *adjust* a variable stimulus so that it seems subjectively equal to a given standard stimulus. In this it differs considerably from the other methods in which the experimenter does the adjustment, usually out of sight or even before the sitting starts, and the subject expresses an opinion on each stimulus but does not alter it. The alternative name of the present method, viz. the Method of Production, suitably emphasises this important psychological point

Mathematically the method presents differences which are mainly due to the fact that in this method almost any value of the variable stimulus can crop up, whereas in the other methods the experimenter customarily keeps to certain steps, so that the results are as it were heaped up at certain points.

The experiment is repeated a large number of times—at least 100—and the arithmetical mean of all the obtained stimulus-values is calculated. The difference between this mean value and the standard stimulus is known as the *crude constant error* e It may be either positive or negative. A measure of scatter is also found.

The crude constant error may be partly due to a space error (the time error cannot occur in this method), partly to other constant

* Thomson, *Brit Journ Psychol* 1912, v. p 205. See also *Brit. Journ. Psychol.* 1914, VI. p. 434, where this method was employed.

† See Francis N Maxfield, "Some Mathematical Aspects of the Binet-Simon Tests," *Journal of Educational Psychology*, 1918, IX. pp 1—12.

conditions*. Let us assume that the experiment is to adjust the length of a variable line until it seems to be equal in length to a standard line. In this case, the standard should be situated to the right in one-half the number of adjustments, and to the left in the other half, either alternately or in haphazard order. The results are tabulated in two columns, I standard to the right, II standard to the left, and the means of these two found separately Half the difference of these means is the space error, while half their sum, less the standard, gives the *residual* constant error.

In order to give as much definiteness as possible to the task of adjustment, the variable should start sometimes shorter than the standard, sometimes (an equal number) longer, and the adjustment be made by lengthening or shortening respectively. Again, the requisite amount of shortening or lengthening should be arranged to be different on different occasions, but alternating with some degree of regularity.

The value of the scatter obtained in this method is from a general point of view a more important result than the value of the constant errors, since it has often been regarded as proportional to the value of the difference threshold as determined by the other two psychophysical methods. The truth is that although there is a certain amount of proportionality between the values, this proportionality is not complete. Under certain conditions the two values vary in opposite directions†. It is hardly necessary to point out that the scatter of the individual points obtained in the Method of Limits has but little relation to that reached by the Average Error Method. They are not entirely unrelated, however.

The closest correspondence of any is that between the distribution of the errors in the present method and that of the judgments "equal" in the Constant Method

(4) THE CONSTANT METHOD

This is generally regarded as the most satisfactory of the psychophysical methods. It can be employed with equal convenience for the determination of absolute thresholds, difference thresholds, equal-appearing sense-distances, and other measurements of psychological importance The different values of the variable stimulus to be employed are fixed once for all at the beginning of the investigation, and are presented to the subject a large number of times (say 100 applications

* Cf E B. Titchener, "Experimental Psychology," *Student's Manual*, II. p. 74· "A constant error is simply an error whose *conditions* are constant; its *amount* may vary, quite considerably, from stage to stage of a long series of experiments "

† This question is connected with the point discussed on p 75 *et seq.*

of each) in irregular order, or in a prearranged order, unknown to the subject, corresponding to certain precautions If an absolute threshold is being determined, the variable is presented alone, if a difference threshold, it is on each occasion preceded, accompanied, or followed by the standard In the latter case, the subject returns the replies *greater*, *uncertain* or *equal*, *less*, with reference either to the standard, or to the variable, or to the first presented stimulus, or on some other prearranged plan. The percentage of each of these three types of answers is determined for each value of the variable used, and recorded (it is also advisable to retain the raw data in the form of the actual answers in the order given)

As one illustration of this method, we shall find it convenient to refer to a series of results obtained by Riecker (*Zeitschrift f. Biologie*, Bd. x.) which has already served in the descriptions of the method given by Muller, Titchener, and Myers. Riecker obtained the following results in an investigation of the "spatial threshold," or threshold of two-point judgments of the skin of the lower eyelid

s (distance between the points of the aesthesiometer in Paris lines)*	0	0·5	1	1·5	2	3	4	5	6
$100p$ (% of two-point judgments)	30	10	14	40	65	80	87	96	100
	30	−20	4	26	25	15	7	9	4

It will be observed that, with two exceptions, the series of percentages follows a general law of increase, the rate of increase itself increasing at first and then diminishing. The numbers in the lowest line are formed by subtracting each percentage from the immediately succeeding one, and show this uniformity more clearly. The two exceptions are at $s = 0$ and $s = 5$. At $s = 0$ the percentage is greater than at the immediately succeeding stimulus. This irregularity is known as an *inversion of the first order*. The cause of it is doubtless to be looked for in the exceptional way in which the stimulus (one point) may have been applied, the pressure and general nature of the contact may have been different from what they were with two-point contact, or some misleading suggestion may have accompanied these particular experiments.

At $s = 5$, the percentage is indeed larger than that for $s = 4$ and smaller than that for $s = 6$, but reference to the differences shows that the increase from 4 to 5 is greater than the increase from 3 to 4, thus breaking the general rule as regards rate of increase. This irregularity is known as an *inversion of the second order*, and being in the present

* A Paris line = 2·25 mms.

case slight, is probably to be explained as due to an insufficiency in the number of applications of the stimulus.

As a second example, in this case one dealing with a difference threshold, we shall take the data for lifted weights which have already been used on p. 51 in connection with the Method of Limits The standard weight was 100 grams and was lifted before each of the seven comparison weights. The judgments given were *lighter* than, *equal* to, or *heavier* than the standard. With Subject I the answers *heavier* were distributed as follows, in 450 trials with each weight·

Comparison weight s	84	88	92	96	100	104	108 grams
Answers *heavier*	1	9	40	100	186	403	423
Proportions p	0022	0200	·0889	2222	4133	8956	·9400

The third row of this table is simply the second row divided by 450*. A figure of the quantities p forms a curve which, following Galton, we

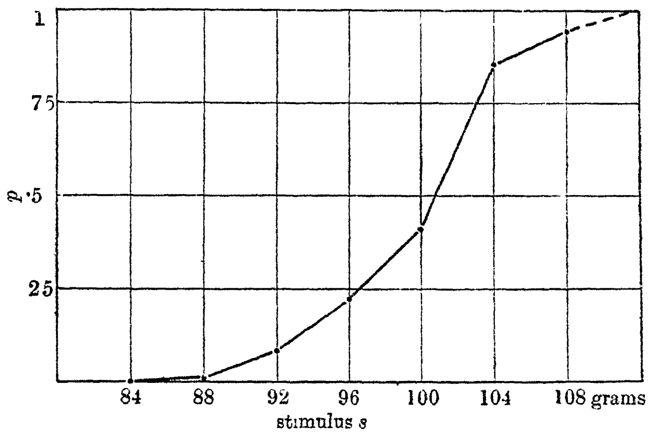


Fig. 7 Urban's Subject I. p = proportion of answers *heavier*

may describe as an ogive. It is shown in Fig 7, where however no attempt is made to draw a smooth curve through the points, which are merely joined by straight lines

Here there are no inversions of either order, and the example is therefore a more convenient one for a first explanation of the various processes of calculation which may be adopted We shall for this reason deal with it first, returning later and considering Riecker's more irregular numbers.

We must commence by defining the limen, and do so in the first

* This figure is correct But (as stated on page 51) Urban only gives 400 values of the just perceptibly heavier points. we are unaware of the reason which led him to discard the other 50.

place as the point corresponding to 50 % of the judgments in question. For example, in our case we define the limen for the response *heavier* as being the point where the subject would give half his answers *heavier*, the other half being *lighter*, *equal*, or whatever other answer is allowed, but not *heavier* Above this point he is more likely to answer *heavier* than not, below it he is more likely not to give this reply.

The limen corresponding to 50 % *heavier* judgments may be calculated in two general ways (A) from the observed values, (B) by finding the best fitting smooth curve, adjusting the observations to this, and then calculating the constants (mean, mode, scatter, etc) from the curve.

A (1). *Linear Interpolation*

The required limen obviously falls somewhere between 100 grams (41·33 % answers *heavier*) and 104 grams (89 56 % answers *heavier*). A very simple way therefore of determining its value is to assume these points joined by a straight line, as in Fig. 7, and find where this line cuts the 50 % line. Arithmetically this means finding a weight which divides the interval from 100 to 104 grams in the same proportion as 50 % divides the interval between 41·33 % and 89·56 %. If T be this weight, we have then

$$\frac{T - 100}{104 - T} = \frac{50 - 41{\cdot}33}{89{\cdot}56 - 50},$$

and with practically no calculation we obtain the value **100·72** grams for the limen.

This method though commendably simple is open to several objections:

(1) It does not employ all the data; it uses two of the percentages only

(2) The assumption that the curve is a straight line at this point is unlikely to be exact

(3) It gives no measure of scatter.

A very simple extension of the above idea has been suggested* which obviates the third and to some extent the first of these objections. This is to calculate the 75 % and the 25 % points by the same simple form of linear interpolation as that employed for the limen itself, that is, on the Fig. 7, to find where the zig-zag ogive crosses the ·25 and ·75 lines. These distances are readily found by a short calculation to be 96·58 and 102·79 grams respectively. Half the interval between

* G H. Thomson, "A Comparison of Psychophysical Methods," *Brit. Journ. Psychol.* 1912, v. p. 210 footnote

them, namely 3·10 grams, is the semi-interquartile range, a rough but very practical measure of scatter.

The fact that the 50 % point is not half-way between the 25 % point and the 75 % point is a rough indication of the skewness of the data it divides the interquartile range in the proportion 4·14 grams below and 2·07 grams above

Moreover, it will be found that in practice the mean of the three values, the 25 %, 50 %, and 75 % points, gives a fairly good approximation to the threshold found by more complicated calculations assuming a symmetrical distribution In the present case this gives

$$\frac{96{\cdot}58 + 100\ 72 + 102{\cdot}79}{3} = 100{\cdot}03 \text{ grams.}$$

A (2). *The Arithmetical Mean** (*Spearman's Formula*)

A value which *can* be calculated by a use of all the data is that of the *mean or average limen.* Before proceeding to consider this, it is important to realise clearly the fact, which G E. Muller† was the first to point out and emphasise, that a limen is a variable magnitude following a law of frequency-distribution There is no fixed limen, only an average limen, a most frequent limen, or a most representative limen. The p's in the table (p 59) represent the relative frequency of limina for stimuli below the corresponding s Thus there were 41·33 % limina below 100 grams, and 22·22 % limina below 96 grams, i.e there were 41·33 − 22·22 or 19·11 % limina between the limits of stimulus-values 96 grams and 100 grams This suggests the plan of plotting a frequency-polygon or histogram for the limina using these differences of the p's. For the present case the table of differences runs

Below	84	grams	1,	or	0022	of the whole
	84—88	,,	8	,,	0178	,,
	88—92	,,	31	,,	0689	,,
	92—96	,,	60	,,	·1333	,,
	96—100	,,	86	,,	·1911	,,
	100—104	,,	217	,,	4823	,,
	104—108	,,	20	,,	0444	,,
Above	108	,,	27‡	,,	0600	,,

* "The Method of Right and Wrong Cases without Gauss's Formulae," *Brit Journ. Psychol.* 1908, II. pp 227—242

† This at least is the view to which Muller himself inclines (*Gesichtspunkte*, p 59), but he deduces his formulae on the assumption, supported by Fechner and Bruns, that the threshold has a single definite value, subject in the course of an experimental determination to variable apparent increase or decrease by random influences which obey the Normal Law of error He points out that both views lead to the same formulae Indeed the distinction is merely a verbal one, in our opinion.

‡ This is not necessarily an inversion of the second order, for it may be spread out to any distance above 108 grams.

and the histogram, or pseudo-histogram* as it is safer to call it, is given in Fig 8

If we now wish to proceed to the calculation of the mean of these limina we must decide where to centre those which lie between say 92 and 96 grams. If we take this centre as the actual midpoint of each interval, then Sheppard has shown† that errors on one side of the mean tend to balance those on the other, and no further correction is required This plan Professor Spearman followed in the first of two alternatives which he discusses, and it is the plan adopted here and referred to when Spearman's Formula is spoken of In his second alternative Professor Spearman proposed another plan with the correctness of which we do

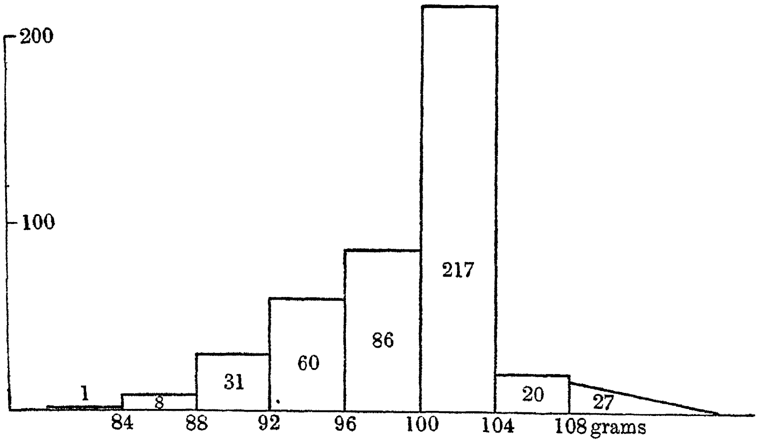


Fig 8 Pseudo-histogram of the 450 limina Urban's Subject I, answers *heavier*

not agree and concerning which Professor Spearman while defending its accuracy informs us in correspondence that it is in his opinion superfluous as the first plan is near enough.

A real difficulty, and one which robs the method of much of the practical utility which it would otherwise have, is *the impossibility of fixing in any unambiguous fashion the centres of the tails of the distribution*, that is those limina which in our example lie below 84 and above 108 grams respectively. This is the same trouble as that referred to on p. 52 in connection with the application of Urban's Formula.

* See later, p 79

† Cf. e g W. F. Sheppard, "On the Calculation of the most probable Values of Frequency Constants," *Proc. Lond Math Soc.* xxix. pp. 353 ff

The simplest thing to do is to assume that at 80 grams the subject would never, and at 112 grams would always, have answered *heavier*, i.e to centre the tails at 82 and at 110 grams respectively We shall do so in this case, although it is clear that at the upper end the tail is probably more spread out than this. We then have the following calculation for the mean or average limen:

1	at	82 =	82
8	„	86 =	688
31	„	90 =	2790
60	„	94 =	5640
86	„	98 =	8428
217	„	102 =	22134
20	„	106 =	2120
27	„	110 =	2970
			44852 total.

Dividing this by the whole number of limina, 450, we obtain **99·67** grams as the threshold by this method. The standard deviation can also be found, $\sigma = 5\ 1$ grams (but see next chapter for a correction to this value, known as *Sheppard's adjustment*).

The summation method of finding the mean, which is explained on p. 19, again enables a saving in arithmetic to be effected in this calculation. Applied directly to the frequencies p it works as follows in the present case, and obviates the actual formation of the pseudo-histogram.

Frequency of answers *heavier* at	84 grams	·0022	
„ „	88 „	·0200	
„ „	92 „	0889	
„ „	96 „	·2222	
„ „	100 „	·4133	
„ „	104 „	·8956	
„ „	108 „	9400	
		2·5822	sum
		4	gram units
		10 3288	
This must be *subtracted from* origin		110	
		99 6712	grams threshold.

The reader must be referred to a study of the summation method on p. 19 to realise why in this case the sum obtained has to be subtracted from the origin.

We can at this point call attention to an instructive comparison which can be drawn between the Method of Limits and the Method of Right and Wrong Cases, as regards the processes of calculation employed

in them, on the basis of Urban's Formula for the former and Spearman's Formula for the latter*.

Urban's Formula, $S(sP)$, shows the threshold as the mean of a number of just perceptible points, which are centred *at* the stimuli used, the quantities P being the differences of successive products of the frequencies of the answers *heavier* (or not *heavier*)

Spearman's Formula, as exemplified in the above calculations, can be written $S(s.dp)$, where dp is the difference of the successive frequencies p, and shows the threshold as the mean of a number of limina which lie *between* the stimuli used.

B. *Methods which fit smooth curves to the data*

The idea underlying these methods is to run a smooth curve through the points of Fig. 7 (the p values) instead of simply joining them by a zigzag, and then to ascertain where this smooth curve, which does not necessarily go exactly through any of the points but smooths off any inequalities, passes the 50 % line. Any suitable curve which happened to occur to one might of course be employed. For example, a parabola of high order can be used†, and the curve $\tan^{-1}\theta$ has also been tried‡. But clearly the whole experiment suggests that an error function of some sort is wanted, and as early as 1860 G. T. Fechner suggested§ that such numbers formed the integral of a Normal Curve of Error.

This idea would naturally occur to anyone accustomed to handling the Normal Curve on considering the pseudo-histogram or table of differences, Fig. 8. The obvious skewness of the diagram would also strike such an observer, it may be noted in passing, but with this point we shall deal in a separate chapter.

The obvious way to fit such a histogram with a Normal Curve is that given in the previous chapter, namely, to find the mean and the standard deviation, and use these constants in the expression for the curve. To this there are however important practical objections, the chief being the difficulty of the undefined tails, to which reference has

* G H Thomson, "A Comparison of Psychophysical Methods," *Brit Journ. Psychol.* 1912, v. p 226

† F. M. Urban, "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," *Archiv f d ges Psychologie,* 1909, xv and xvi. pp 335—355.

‡ Urban, *loc cit.* p. 393 *et seq*

§ G. T Fechner, *Elemente der Psychophysik,* 1860.

already been made in discussing Spearman's plan of finding the mean. This difficulty is still more acute when we attempt to find the standard deviation This tail difficulty does not occur in the plan adopted by Muller*, which fits the Normal *Integral* direct to the p values, not troubling at all about the differences forming the histogram. In other words, Muller's process fits a curve to Fig 7, not to Fig. 8, and is therefore more direct, in addition to the advantage it has of avoiding the tail problem.

Before proceeding to the explanation of Muller's process, it is necessary to notice that he employed a slightly different form of the equation to a Normal Curve from that used in the previous chapter. The latter is the form now in general use among biometricians, and it seems desirable that it should also be used by psychometricians, who otherwise would be hindered from the direct application to their work of the mathematical improvements made by the biometric school, and especially would find their use of the valuable tables published by Professor Pearson much hampered In the actual description of Muller's work however it is better for the present to keep to his notation in this respect. The form of the Normal Curve used by him was

$$y = \frac{h}{\sqrt{(\pi)}} e^{-h^2(s-T)^2}.$$

In this s is the variable stimulus, T is the average limen, and therefore also, since the curve is symmetrical, the median and the mode So far the notation agrees with that used in the expression employed for the Normal Curve in the previous chapter, namely

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(s-T)^2/2\sigma^2},$$

but a further comparison of the two shows that for his second constant Muller has a quantity h, which is connected with the standard deviation σ of the biometric formula by the relationship

$$h^2 = \frac{1}{2\sigma^2}.$$

When the standard deviation is large, therefore, h is small. It is a *measure of precision*, not of scatter.

The assumption is now made that the relationship between the

* G E Muller, "Ueber die Maassbestimmungen des Ortssinnes der Haut mittels der Methode der richtigen und falschen Falle," *Pflüger's Archiv fur die ges Physiologie*, 1879, XIX. pp. 191—235, especially par 5 *et seq.*: also *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, Wiesbaden, 1904, par 11, where the classical description of this method will be found.

stimulus s, and the frequency p with which the answer *heavier* is returned, is given by the equation

$$p - \int_{-\infty}^{s-T} \frac{h}{\sqrt{(\pi)}} e^{-h^2(s-T)^2} ds = 0 \quad \ldots (1).$$

That is to say, it is assumed that the successively increasing percentages of answers *heavier* correspond to the increasing area of the portion of a Normal Curve which is shaded in Fig 9, as the point s moves to the right in that figure

To obtain this equation in a more convenient form for our purpose, write

$$h(s - T) = t \quad \ldots\ldots(2).$$

This corresponds to measuring the stimuli in a special unit, and is the

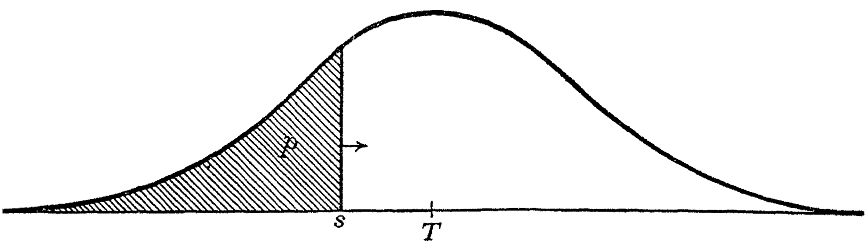


Fig 9 To illustrate Muller's process

same device as that used in another connection later by Galton. The equation then becomes

$$p - \frac{1}{\sqrt{(\pi)}} \int_{-\infty}^{h(s-T)} e^{-t^2} dt = 0 \quad \ldots\ldots(3).$$

By inserting in this equation the corresponding values of p and s from the table of data on p. 59, we obtain seven equations for two unknowns h and T, and as these equations are slightly inconsistent with one another we have to decide how to calculate the most probable values of h and T. No pair of values will exactly satisfy all seven equations Instead of coming to zero they leave small *residuals* v. Muller adopted the Method of Least Squares, an account of which will be found on p. 44 in Chapter II.

In passing a note must be made of the fact that Müller assumed tacitly that these observation equations, being each based on the same number of experiments, are of equal importance or "weight*." We shall allow this assumption to pass for the present but shall return to it later.

Unfortunately, the equations are very far from being simple and

* There is unfortunately a possibility of ambiguity here owing to the fact that *weights* are used as stimuli.

linear as in the example on pages 44 and 45 To avoid this difficulty, we look up in tables of the Probability Integral* those values of

$$\gamma = h(s - T) \qquad \dots \quad (4)$$

which correspond exactly to our values of p

These equations are not yet linear in T and h, though much simpler. If however we write

$$c = hT,$$

they become

$$\gamma - hs + c = 0 \qquad \dots\dots \quad (5)\dagger$$

and are now linear in h and c If we now insert any pair of values h and c into these seven (or in general, n) equations, these also will leave residuals u, different from those considered above in connection with the equations (3). If we were now to proceed to make $S(u^2)$ a minimum, this would not effect our purpose. It is $S(v^2)$ we wish to make a minimum, not $S(u^2)$ If however we can find multipliers or "weights" M such that each

$$Mu^2 = v^2,$$

then we can make

$$S(Mu^2)$$

a minimum. That is, we can apply Least Squares to the equations (5) weighted with certain artificial weights (in addition to any weights which may possibly be necessary by reason of there being different numbers of experiments at the different stimulus-values). The use of this device of artificial weights to overcome the complexity due to the non-linear equations is Muller's particular credit in this connection

Clearly the residuals v, which may be regarded as errors in p, are connected with the residuals u, which may be regarded as errors in γ, by the equation

$$\frac{1}{\sqrt{(\pi)}} e^{-\gamma^2} u = v$$

from equations (3) and (4). Therefore

$$M = e^{-2\gamma^2}/\pi.$$

Herein we can omit the π since it is only the relative values of the Muller weights which are of importance These weights are, by reason of the improved weights to be shortly described, of only historical interest. The condition that $S(v^2)$ should be a minimum has now

* The table of the Probability Integral commonly used by psychologists is that known as Fechner's Fundamental Table, which is given in Appendix I. It is desirable that psychologists should make the slight changes in notation necessary to enable them to use better and more generally accessible tables For example, the first table in Pearson's *Tables for Biometricians and Statisticians* is, except for a factor $\sqrt{2}$, identical in significance with Fechner's, and gives more values, to more decimal places

† Note that this, like (3), represents a *set* of equations, each of this form. In our example there are seven such equations.

become that $S(Mu^2)$ should be a minimum. With this substitution, the equations (5) give the two Normal Equations

$$\left.\begin{aligned} S(Ms\gamma) - S(Ms^2)\,h + S(Ms)\,c = 0 \\ S(M\gamma) + S(Ms)\,h - S(M)\,c = 0 \end{aligned}\right\} \quad \ldots\ldots(6).$$

Thence we have

$$\left.\begin{aligned} c &= \frac{S(Ms)\,S(Ms\gamma) - S(M\gamma)\,S(Ms^2)}{S(M)\,S(Ms^2) - S^2(Ms)} \\ h &= \frac{S(M)\,S(Ms\gamma) - S(Ms)\,S(M\gamma)}{S(M)\,S(Ms^2) - S^2(Ms)} \\ T &= c/h \end{aligned}\right\} \quad \ldots\ldots(7).$$

The use of these formulae is best explained by an example. Before giving one, however, it is well to describe a modification in the weights M which was introduced in 1909 by Professor F. M Urban*, as it is with Urban's weights, and not with Muller's, that we shall actually work

These alterations in the Muller weights, or rather additions to Muller's weights, which Professor Urban made, arise from the notion of the probability of a certain judgment, with which we are already familiar The analogy is, as it were, between extracting the answers *heavier* or *lighter* from a subject, and extracting black or white balls from a bag containing a mixture of these colours. Compare, for example, the two statements.

(1) From a bag containing black balls and white balls, 450 drawings are made, one at a time, the ball being returned each time before the next drawing is made. 403 black balls are observed out of the 450.

(2) A subject on performing a certain experiment with lifted weights sometimes gives the answer *heavier*, sometimes some other answer. On one occasion, when the weights were 100 grams standard and 104 grams unknown, this experiment was repeated 450 times, and the answer *heavier* was obtained 403 times out of the 450.

Now if p is the observed proportion (here 403/450) of black balls in a bag, then the probable error of p is known to vary with $\sqrt{p(1-p)}$, or $\sqrt{pq}$†. With the same sized sample, a result $p = \cdot 5$ has a larger probable error than a result $p = \cdot 8$ say. If anything similar holds, as the analogy suggests, for the psychometric experiment, then the seven, or n,

* "Die psychophysischen Massmethoden," *Archiv f. d. ges. Psychol.* 1909, XV and XVI. p. 357 *et seq.*

† Cf. p. 32, Standard Deviation of the Binomial Expansion Really the *true* values of p and q should be used, but this is the best we can do And further, the expression probable error ceases to have an accurate meaning when p is too close to zero or unity and the distribution is in consequence very skew. But these refinements do not affect the argument except in detail.

equations (5) are not equally reliable, even though based on the same number, 450, of experiments each. In addition to the Muller weights M they need other weights $1/4pq$ to allow for this new variation in reliability.

These weights, it will be observed, arise from the fact that drawing balls singly from a bag in this way gives rise to a binomial distribution, and the standard deviation of such a distribution is, as is shown on p. 32 in the previous chapter, equal to $\sqrt{pq}$. The combined weights $M/4pq$ are known as Urban's weights, and are also given in a table in Appendix I. Professor Urban discusses the matter at some length in his already cited article, and a discussion will also be found in Wirth's *Psychophysik* (Leipzig, 1912), where on p. 151 the actual scatter of various p's is given in a diagram.

The fact that in the above we have taken the weights as proportional inversely to the *square* of the probable error, pq, need cause the reader no trouble, for it is the same phenomenon as the fact that the accuracy of a set of readings increases with the square root of the number of readings made, as follows from p. 24. The weight of an observation equation in an ordinary sense is simply the number of times the observation has been repeated, that is it is proportional to n the number of observations, which as we have just said is proportional inversely to the square of the probable error.

Using the symbol W for the combined Urban and Muller weights which are given in tabular form in Appendix I, we have to replace M by W in the equations (6). These equations we shall now illustrate by giving at some length the calculations indicated by them in the case of Urban's Subject I, answers *heavier*.

We first look up in Fechner's Fundamental Table*, i.e. a table of the probability integral, those values of γ which correspond to the observed values of p. These are given in the third column of the adjoining table. The fourth column of that table gives the values of the Urban-Muller weights W, found from the table in the Appendix.

Urban's Subject I, Heavier *answers*

Values of γ and W for substitution in the normal equations

Stimulus s grams	p	γ	W	Ws
84	·0022	−2·0150	0·025	2·10
88	·0200	−1·4520	0·187	16·46
92	·0889	−0·9528	0·502	46·18
96	·2222	−0·5408	0·806	77·38
100	·4133	−0·1549	0·982	98·20
104	·8956	0·8888	0·551	57·30
108	·9400	1·0993	0·396	42·77
			3·449	340·39

* See Appendix I

From this table $S(W)$ is at once obtained, and the other quantities appearing in equations (6) are easily formed, though the arithmetic is laborious. The work for $S(Ws)$ is given in the last column of the above table, and the other quantities are obtained similarly. They prove to be

$$\begin{aligned} S(W) &= 3{\cdot}449, \\ S(Ws) &= 340\ 39, \\ S(Ws\gamma) &= -31{\cdot}223, \\ S(W\gamma) &= -0{\cdot}463, \\ S(Ws^2) &= 33700{\cdot}1. \end{aligned}$$

The equations (7), reading W instead of M, then give

$$\text{Threshold } T = \mathbf{99\ 68} \text{ grams,}$$

$$\text{Precision } h = 0{\cdot}136113.$$

The standard deviation corresponding to this precision is

$$\sigma = \frac{1}{\sqrt{2}.h} = 5{\cdot}2 \text{ grams,}$$

and $\cdot 6745\sigma = 3{\cdot}5$ grams.

The fact that 99·68, which is thus found for the threshold of answers *heavier*, is actually smaller than the standard 100 grams than which it is judged heavier, may cause confusion if it is not at once explained that this value involves a time error

It will be seen that the Constant Process as described above and as illustrated by this example, involves a great deal of arithmetical work, so much so indeed that it is certain never to be used except in some special cases, unless plans for easing this labour be adopted. Something can be done by using the arithmetical short-cuts explained in the preceding chapter, and Crelle's Calculating Tables, or better still a calculating machine, makes the work practicable. But the best device for reducing the arithmetical work involved in the Constant Process is that adopted by Professor Urban in publishing his tables for this method. These tables are given in Appendix I, and do away with the necessity for Fechner's Fundamental Table and the Table of the Muller-Urban weights. They assume that exactly 100 experiments have been made at each stimulus-value, but of course if other numbers of experiments have been made, the p's can be approximated to by two significant figures. Their use will be readily grasped from the worked example on p. 73 below: and that example is also employed in Appendix I to explain Rich's useful Checking Table.

We give next, as models, the calculations by all the above methods for Riecker's data, assuming however that no two-point judgments were given at zero stimulus-distance The data are given on p 58 and it should be noted that the stimuli are not equidistant, and that therefore summation methods cannot be used to lessen the arithmetical work.

Riecker's Data

(1) Limiting Process. By Urban's Formula

(a) *Just Perceptibly-two Points*

s	p	q	q products	P	Ps	Ps^2
0	00	1 00	1 0000	0000	0000	0000
0 5	10	90	9000	1000	0500	0250
1	14	86	7740	1260	1260	1260
1 5	40	60	4644	3096	4644	6966
2	65	35	1625	3019	6038	1 2076
3	80	20	0325	1300	3900	1 1700
4	87	13	0042	0283	1132	4528
5	96	04	0002	0040	0200	1000
6	1 00	00	0000	0002	0012	0072
			Sums	1 0000	1 7686	3 7852

$$T = \text{sum of } Ps = 1{\cdot}7686 \text{ Paris Lines}$$

Standard deviation of the just perceptibly-two points, squared,

$$= \text{sum of } Ps^2, \text{ less } T^2$$
$$= 3{\cdot}7852 - 3{\cdot}1279 = {\cdot}6573,$$

whence standard deviation = ·81 Paris Lines.

(b) *Just Imperceptibly-two Points*

s	p	p products	P'	$P's$	$P's^2$
0	00	0000	0024	0000	0000
0 5	·10	·0024	0219	0109	0055
1	14	0243	·1494	·1494	1494
1 5	40	·1737	·2606	3909	5863
2	·65	4343	2339	4678	·9356
3	80	·6682	·1670	5010	1 5030
4	87	·8352	·1248	4992	1 9968
5	96	9600	0400	·2000	1 0000
6	1 00	1 0000	0000	0000	0000
		Sums	1 0000	2·2192	6 1766

whence

$$T' = 2\ 2192 \text{ Paris Lines},$$
$$\sigma' = 1{\cdot}12 \text{ Paris Lines},$$
$$(T + T')/2 = \mathbf{1{\cdot}99} \text{ Paris Lines},$$

Mean of the two σ's, **0·97** Paris Lines.

(2) Linear Interpolation

for 75 %, 50 % and 25 % points.

$$\frac{80-75}{75-65}=\frac{3-Q_2}{Q_2-2},\quad Q_2 = 2\ 67,$$

$$\frac{65-50}{50-40}=\frac{2-T}{T-1{\cdot}5},\quad T = 1{\cdot}70,$$

$$\frac{40-25}{25-14}=\frac{1{\cdot}5-Q_1}{Q_1-1},\quad Q_1 = 1{\cdot}21,$$

$$\left.\begin{array}{l} Q_2 - T \ = 0{\cdot}97 \\ T - Q_1 = 0{\cdot}49 \end{array}\right\}\text{(skew).}$$

Interquartile Range = 1·46.
Semi-interquartile range = 0·73.
$(Q_1 + T + Q_2)/3 =$ **1·86** Paris Lines.
Rough value for σ, 0·73/0·6745 = **1·08** Paris Lines.

(3) Arithmetical Mean (Spearman's Formula)

s	p	dp	centre s'	$dp \times s'$	$dp \times s'^2$
0	00	10	25	0250	0062
0 5	·10	·04	75	·0300	·0225
1	·14	26	1 25	·3250	4063
1 5	·40	25	1 75	4375	·7656
2	·65	15	2 5	·3750	9375
3	·80	07	3 5	2450	·8575
4	·87	09	4 5	·4050	1 8225
5	96	04	5 5	·2200	1·2100
6	1 00	·00			
		1 00	Sums	2 0625	6 0281

Threshold $T =$ **2·0625** Paris Lines.
Square of standard deviation = 6·0281 − T^2 = 1·7742,
Standard deviation = **1·33** Paris Lines*.

* Without Sheppard's correction, for which see p. 84.

(4) CONSTANT PROCESS

using Urban's Tables. (Appendix I, p 194)

s	working s	p	W	γW	sW	s^2W	$s\gamma W$
0	−6	00	0000	0000	0000	0000	0000
0 5	−5	10	5376	− 4871	−2 6878	13 4388	2 4356
1	−4	14	6463	− 4937	−2 5853	10 3413	1 9749
1·5	−3	40	9768	− 1750	−2 9306	8 7916	5252
2	−2	65	9473	2581	−1 8945	3 7890	− 5163
3	0	80	·7695	4579	0000	0000	0000
4	2	87	6215	4950	1 2430	2 4860	9900
5	4	96	3036	3759	1 2146	4 8582	1 5036
6	6	1 00	0000	0000	0000	0000	0000
		Sums	4 8026	1 5869	2 4576	43 7049	7 4293
				−1·1558	−10 0982		− 5163
				4311	−7·6406		6 9130

$$T = \frac{-7\ 64 \times 6\ 91 - \cdot 431 \times 43\ 7}{4\cdot 80 \times 6\cdot 91 + \cdot 431 \times 7\ 64} = -1\cdot 96$$

in working units from the working origin

$$= 3 - \frac{1\cdot 96}{2} = \mathbf{2\cdot 02} \text{ Paris Lines,}$$

$$h = \frac{4\cdot 80 \times 6\cdot 91 + \cdot 431 \times 7\cdot 64}{4\ 80 \times 43\ 7 - 7\cdot 64^2} = \cdot 241 \text{ in working units,}$$

whence $\sigma = 1/(\sqrt{2}h) = 2\cdot 93$ working units

$= \mathbf{1\cdot 46}$ Paris Lines.

Titchener, using the Constant Process with Muller weights alone (the above is with the Muller-Urban weights), obtained

$T = 1\cdot 88$ Paris Lines,

$h = 0\cdot 49$,

whence $\sigma = 1\cdot 44$ Paris Lines.

Summarising in one table we have

Riecker's Data calculated by different Processes

Process	Threshold T	Scatter σ
Limiting Process	1 99	0 97
Linear Interpolation	1 86	1 08
Arithmetical Mean	2 06	1 33
Constant Process	2 02	1·46

The Probable Error of the Thresholds calculated in these different ways.

The values of the standard deviation given in the table immediately

above refer of course to the variation of the individual limina of which the threshold T is a central measure If there are n of these individual limina, then in an ordinary way the arithmetical mean of these would have a standard deviation of $\sigma/\sqrt{n}$. The standard deviations of the above values of T are indeed of this *order of magnitude*, say in the first case

$$0{\cdot}97/\sqrt{(100)} = {\cdot}097,$$

but to avoid misconception ought to be regarded as distinctly larger than this. This arises from various causes which cannot here be gone into In the case of the Limiting Process there is the dependence upon the particular choice of stimuli, in the case of Spearman's Arithmetical Mean formula there is the uncertainty about the centring of the "tails" of the distribution, etc.

Professor F. M. Urban, some twelve years ago*, brought forward reasons which in his opinion showed that the Method of Limits was much more exact than the Constant Method but his mathematics contained certain errors†. The various processes do not differ very widely in this respect if each is used in circumstances favourable to itself, but on the whole the Constant Process is most reliable, and Linear Interpolation least.

In conclusion, we may venture to express a cautious opinion on the choice of a process of calculation from among those given. Frequently of course there is no choice, for the conditions of experiment fix the matter for us. For example, if the just perceptible points have been recorded but not the frequencies p of answers of a certain kind at each stimulus-value, then the direct Limiting Process must be employed. We will suppose however that the fullest records have been taken

If the points at which $p = 0$ and $p = 1$, that is the points where only one kind of answer is given, are known or are very nearly approached, the Arithmetical Mean of Spearman is in our opinion best

If however, as is frequently the case, these points are unknown, then the simple linear interpolation for the 25 %, 50 %, and 75 % points is a good plan from the point of view of simplicity and is often of sufficient accuracy.

If the accuracy of the data justifies the use of the full Constant Process, then Urban's Tables lighten the work enormously, and give

* "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," *Archiv f d. ges. Psychol* 1909, XV and XVI pp 261—415

† G H. Thomson, "Note on the Probable Error of Urban's Formula for the Method of Just Perceptible Differences," *Brit Journ. Psychol.* 1913, VI. p 217 and "The Accuracy of the Phi-gamma Process," *ibid.* 1914, VII. p 44

perfect accuracy if exactly 100 experiments have been made at each stimulus The great advantage of the Constant Process lies in the fact that the "tail" difficulty does not arise. Experiments in which it is on psychological grounds inadvisable to employ extreme stimuli can only be handled by this process

But it is not worth while applying it unless the calculator is assured that the experimental accuracy justifies it, and unless he has convinced himself, by methods to be described in the next chapter, that the distribution is not significantly skew, and the data not heterogeneous Very few collections of psychophysical data are worth the accuracy of the Constant Process, which is however undoubtedly the best theoretically for symmetrical distributions.

(5) DIFFERENCE THRESHOLDS AND THE PROBABILITY OF A JUDGMENT OF A CERTAIN CATEGORY

A question closely bound up with the mathematics of the psychophysical methods is that of the best measure of a subject's sensitivity to differences of stimulus-value.

To fix ideas, we shall use the case already discussed of the difference threshold for lifted weights. When a sufficient number of judgments has been collected, the three categories *lighter*. *equal* or *undecided*, and *heavier*, are found to occur with varying frequency with the different comparison weights. The difference threshold is then decided by the positions of the points T and T' where the descending *lighter* and ascending *heavier* curves cross the halfway line (see Fig 7, p 59) The distance $(T - T')/2$ or some closely similar quantity is what is called the difference threshold, and is commonly used in comparing the sensitivity of different subjects. The smaller $T - T'$, the more sensitive the subject is said to be.

This distance however depends entirely on the subject's readiness to give the answer *undecided*. It measures therefore rather a moral characteristic than a physical sensitivity, and varies very much with the instructions given to the subject The moral character of the measure $T - T'$ is above all seen from the fact that any subject who wishes may reduce it to zero, whatever may be his actual sensitivity to differences of weight, simply by determining that he will never give the answer *undecided*

There is however another measure which has been used. This can be most conveniently described by considering first a case in which a subject gives no undecided answers. In such a case, the thresholds T and T'

have come together and on the previous plan the subject's sensitivity would be considered as infinite, and all subjects giving no *undecided* answers would have the same infinite sensitivity whereas clearly the subject's sensitivity is connected with the rapidity with which the curves pass from 0 to 1 or *vice versa*, and two subjects may differ very much in this respect even although they both give no *undecided* answers. Under these circumstances a measure which has been used is the distance $Q - Q'$, the horizontal distance between the crossing of the ·25 and ·75 lines (see Fig. 7, p. 59) Under another guise it was used by Fechner also for the cases where undecided answers *were* given. In such cases he reduced the three curves to two by sharing the *undecided* answers between *heavier* and *lighter*.

This measure has the advantage that the subject cannot increase his apparent sensitivity at will, as was the case with the "threshold" measure. $Q - Q'$ is the interquartile range of the point of subjective equality, represented by the crossing of the heavier and lighter curves. It and the difference threshold measure distinctly different things, and subjects placed in order of merit by the one will be found in a different order by the other.

The points here raised seem to suggest an extension of Urban's idea of the probability of a judgment, which compares the giving of the judgments, *heavier*, *undecided* or *lighter*, with drawing a ball from an urn containing say red, white, and blue balls, and ascertaining its colour. For each stimulus the urn is supposed to contain different porportions of the coloured balls.

In place of this is suggested the following. For each stimulus imagine an urn containing an infinite number of balls some black and some white, in a proportion varying in some way with the stimulus. A judgment may then be compared with taking not one but a handful of balls from the urn, the *kind* of judgment depending upon the proportion of black balls in the handful.

From this point of view, the standard weight, the variable weight, and the physiological make up of the subject decide the proportion of black balls in the urn: but the decision as to what proportion is to be called *heavier*, what *undecided*, and what *lighter*, depends upon a conscious act of the subject*.

* See Thomson, *Psychol Rev* 1920, XXVII. p 300 In a very interesting reply by Boring (*ibid* p. 440) to which Thomson regrets that lack of time has prevented an answer, but with which he in the main agrees, it is suggested that "the subject must be both instructed and trained to maintain a constant attitude throughout the experiment," in which case "he will not give doubtful judgments" References are given to George, *Amer. Journ Psychol* 1917, XXVIII p 1, to Boring, *ibid* p 465, and to others.

CHAPTER IV

SKEWNESS AND HETEROGENEITY IN PSYCHOPHYSICAL DATA

Obvious skewness of many psychophysical curves—Pearson's test for goodness of fit applied to the method of average error—Applied to the method of right and wrong cases—Skew curves in homogeneous material—The summation method of finding moments—Calculation of a skew curve—Analysis into two normal curves—Conclusions

(1) OBVIOUS SKEWNESS OF MANY PSYCHOPHYSICAL CURVES

To anyone accustomed to handling distribution data, a most striking point about the results of many psychophysical experiments is the obvious skewness of much of the data Both the examples used extensively in the previous chapter show this very strongly as can be seen from an inspection of the data either in numerical or diagrammatic form, and also from the various values of the threshold T found by the different processes of calculation. for these differences arise largely from the fact that the distribution is not normal.

Taking the best of the processes, the Constant Process, it is of interest to see how closely the curve which it gives fits the original data. A method of thus estimating the goodness of fit of curves has been given by Professor Karl Pearson His method is perfectly general, and applicable to all classes of curves*, but it has been most fully worked out for the fitting of bell-curves to histograms Our problem is not of this nature, though it might appear to be so, for the pseudo-histogram (Fig. 8) which can be formed from the frequencies p differs essentially from a real histogram. Since in psychophysics it may often be necessary to fit curves to real histograms, for example those obtained in the Method of Average Error, we shall first explain Pearson's Goodness of Fit Test for this case, using the bisection data of Chapter II for the purpose (see pp. 15 and 42 and Fig 6).

(2) PEARSON'S TEST FOR GOODNESS OF FIT APPLIED TO THE METHOD OF AVERAGE ERROR

The bisection data had a mean of 60·13 mms and a standard deviation of 1·38 mms. With these values, using Sheppard's Tables, we draw the smooth curve shown in Fig 6 Now it is important at the outset to

* *Phil. Mag* July 1900, Fifth series, L. pp. 157—175. *Phil. Mag.* April 1916, Sixth series, XXXI. pp. 369—378.

realise that whether that curve is a good or bad fit to the data depends on the number of observations made The number in this case was only 29, and it will presently be shown that the curve is a very good fit But had the number of observations been 2900 it would have been a bad fit, for with such a number of observations the histogram ought to have modelled itself more closely to the curve

In order to apply Pearson's test, we must find the theoretical histogram for comparison with the observed histogram That is, we must find the areas of the slabs of the curve in Fig 6 which replace the rectangles (One of these slabs is cross-hatched in that figure to explain more clearly what is here meant) This is most easily done from Sheppard's Tables by calculating the areas of the smooth curve from $-\infty$ up to each dividing ordinate in turn, and taking the differences of these numbers, as is done in the following table The quantity $\frac{1}{2}(1+\alpha)$ in Sheppard's Tables is the area of a Normal Curve, of unit total area, up to the ordinate x/σ. For negative x's it has to be subtracted from unity.

Calculation of the Theoretical for comparison with the Observed Histogram of the Bisection Data

x mms	$x' = x - 60\ 13$	Sheppard's $x = x'/1\ 38$*	Sheppard's $\frac{1}{2}(1+\alpha)$	Multiplied by 29*	Differences
					0·09
63 95	3 82	2 77	·997	28 91	0 52
62 95	2 82	2 04	979	28 39	2 09
61 95	1 82	1 32	907	26 30	5 36
60 95	0 82	0 59	·722	20 94	7 95
59 95	−0 18	−0 13	·448	12 99	7 33
58 95	−1 18	−0 86	·195	5 66	4 01
57 95	−2 18	−1 58	·057	1·65	1 36
56 95	−3 18	−2 31	010	0 29	0 29
					29 00

The actual and theoretical histograms are then compared in the next table. Pearson's method, the theory of which cannot be given here, then forms the quantity

$$\chi^2 = \text{Sum}\left(\frac{\text{square of differences of theoretical and observed frequencies}}{\text{theoretical frequency}}\right).$$

With this value χ^2, and n' the number of cells in the histogram, Table XII in Pearson's *Tables* is then entered and a value of P found.

* These can be calculated at one opening of Crelle's Tables, or by one setting of a slide rule.

n' is here 9 counting in the two tail cells P is then the probability that the observed *or a worse* distribution will be obtained (assuming the theoretical distribution) in a sample of the size taken, here 29.

Calculation of Goodness of Fit of Normal Curve to the Bisection Data

mms	Actual observations m	Theoretical m'	$(m-m')^2/m'$
Above 63 9	0	0 09	090
63—63 9	1	0 52	443
62—62 9	2	2 09	004
61—61 9	6	5 36	076
60—60 9	6	7 95	478
59—59 9	8	7 33	061
58—58 9	5	4 01	244
57—57 9	1	1 36	095
Below 57	0	0 29	·290
	29	29 00	1 781 $=\chi^2$

No. of cells $n' = 9$

From Elderton's Tables, No. XII in Pearson's *Tables*,

$$P = 0{\cdot}998 \text{ for } \chi^2 = 1,$$
$$0{\cdot}981 \text{ for } \chi^2 = 2,$$

therefore $P = 0{\cdot}99$ approximately.

In our case we find $P = 0\ 99$, i.e 99 samples of this normal distribution out of 100 would give no better a fit than the present The Normal Curve is therefore a perfectly satisfactory theory for these 29 observations

(3) PEARSON'S TEST OF GOODNESS OF FIT APPLIED TO THE METHOD OF RIGHT AND WRONG CASES

The above plan of testing goodness of fit cannot however be applied to the pseudo-histogram of Fig 8*. The reasons why this is so cannot be here gone into in detail, but they are based upon the following differences between the two cases. In a real histogram, if any one of the cells is larger than it ought to be, then any other must have a tendency to be smaller than it ought to be. There is a strong negative correlation between the numbers in the cells, a correlation, that is, from trial to trial. In the psychometric pseudo-histogram, however, formed from the proportions p, this is otherwise, because the p's are measured quite separately from one another. In a real histogram the numbers in

* G. H. Thomson, "The Criterion of Goodness of Fit of Psychophysical Curves," *Biometrika*, 1919, XII. pp. 216—230.

each cell are necessarily positive quantities. In the psychometric pseudo-histogram they may be negative, if the p's do not rise steadily.

Psychometrical data of the kind here considered, in fact, as has already been pointed out, are not really in histogram form. Although a kind of histogram can be deduced from them, it is only by making certain assumptions, and the intercorrelations of the cells of this artificial histogram are different from the intercorrelations of a naturally observed histogram Under these circumstances we must in applying the goodness of fit test turn to the directly observed quantities p. These are compared with the values calculated from the Constant Process in the following table, the remaining columns of which are explained below:

Calculation of Goodness of Fit of a Constant Process Ogive

Urban's Subject I, answers *heavier*

Grams s	Observed p	Calculated p'	$p-p'$	$(p-p')^2/p'q'$
84	0022	0013	0009	001
88	0200	0122	0078	005
92	0889	0697	0192	·006
96	·2222	·2394	− 0172	002
100	·4133	·5246	−·1113	050
104	8956	·7972	0984	·060
108	9400	·9454	− 0054	001
				$125 = S\{(p-p')^2/p'q'\}$

$$\chi^2 = 450 \times \cdot 125 = 56{\cdot}25,$$
$$n' = 8 \text{ (one more than the number of stimuli)},$$
$$P = \cdot 0000005 \text{ from Pearson's } \textit{Tables}.$$

To these quantities p the principles underlying Pearson's test can be applied direct. They are indeed the same principles already used in Chapter II when we were discussing curve fitting. We have n quantities p which are independently measured, and n quantities p' which are theoretically given. The variations of p from p' are binomial in form, that is approximately normal. If we look upon the judgment *heavier*, as suggested in an earlier paragraph, as being comparable with drawing black balls out of a bag containing black balls and white balls in the proportions p' and $1-p'$, then the probable error of p is

$$\cdot 6745\sqrt{\{p'(1-p')/k\}},$$

k being the number of judgments of which pk are of the category *heavier*.

For the chances of obtaining 0, 1, 2, ..., $k-1$, or k black balls in

a drawing of k are given by the terms of the binomial $(p' + q')^k$, q' being $1 - p'$, that is, the chances of obtaining

$$p = 0/k,\ 1/k,\ ..\ (k-1)/k,\ k/k \text{ or unity}$$

The standard deviation of the above binomial is $\sqrt{(kp'q')}$, and the standard deviation of p therefore $\frac{1}{k}\sqrt{(kp'q')}$, or $\sqrt{(p'q'/k)}$. The probability of an error $p - p'$ is therefore

$$\frac{\sqrt{k}}{\sqrt{(2\pi p'q')}}\, e^{-\frac{k}{2p'q'}(p-p')^2}$$

The probability of the whole set of observed values $p_1, p_2, \ldots p_n$ occurring is the product of n such factors, and is of the form

$$z = z_0 e^{-\frac{1}{2}\chi^2};$$

where

$$\chi^2 = S\left\{k\frac{(p-p')^2}{p'q'}\right\},$$

or if k is the same at each stimulus,

$$\chi^2 = kS\frac{(p-p')^2}{p'q'}.$$

The remaining columns in the above table calculate this quantity*, which is the same as Pearson's χ^2, under our special circumstances. For n' we have to use $n + 1$, because all the n values of p can vary separately, whereas in a real histogram the number of variables is one less than the number of cells. We thus reach the value $P = \cdot 0000005$, that is, the curve is an incredibly bad fit to the data, which cannot possibly be regarded as differing from a normal distribution by sampling errors alone.

(4) SKEW CURVES IN HOMOGENEOUS MATERIAL

There are two immediate hypotheses which present themselves to explain this bad fit of the normal curve, (*a*) that the material is not homogeneous, the conditions of experiment not having remained the same throughout, (*b*) that the material is homogeneous, but the underlying factors which cause the distribution of errors or deviations are not independent, but correlated, as in the system of generalised probability curves next to be described. A word of caution may first be given, for, as we have already suggested, the identification, or even the approximation, of graphical representations of "percentage" judgments to true frequency-distributions is of extremely doubtful validity.

* A table in Appendix I, giving values of $1/(pq)$, lightens the work considerably.

The general theory of curve fitting has been worked out in great detail by Professor Karl Pearson*. A good account of his method is given in a book by W. Palin Elderton, *Frequency Curves and Correlation*, C. and E Layton, London, 1906, to which the mathematical reader is referred for fairly complete theoretical and practical information

Frequency curves of data not involving a mixture of species tend to commence at zero, rise to a maximum, and then fall either at the same or at a different rate. There is often high contact at one or both ends of the distribution. An equation of the general form

$$\frac{dy}{dx} = \frac{(x+a)\,y}{f(x)}$$

satisfies both these conditions, since if $y = 0$, $dy/dx = 0$ (high contact), and if $x = -a$, $dy/dx = 0$ (maximum, for a maximum, again, the second differential coefficient must be negative). Expanding $f(x)$ by Maclaurin's Theorem, we have

$$\frac{dy}{dx} = \frac{(x+a)\,y}{c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots}.$$

(1) Putting $c_1 = c_2 = c_3 = \ldots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{c_0},$$

which is the Gaussian or normal curve† It fits the symmetrical binomial $(\frac{1}{2} + \frac{1}{2})^n$, e.g. in coin tossing, where the chances for and against are equal ($p = q$), and the contributory causes are independent of one another.

(2) Putting $c_2 = c_3 = \ldots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{c_0 + c_1 x},$$

which represents a class of curves varying from the Gaussian curve to the J-curve It fits the asymmetrical or "point" binomial $(p+q)^n$, e g. in teetotum spinning or dice throwing, where the chances for and against are not equal, $p \neq q$, but the contributory causes are still independent of one another

* Karl Pearson, "Skew Variation in Homogeneous Material," *Phil Trans* 1895, CLXXXVI A, pp 343 ff ; "On the Systematic Fitting of Curves to Observations and Measurements," *Biometrika,* I. pp. 265 ff and II. pp 1 ff, 1901—3; "On the Curves which are most suitable for describing the frequency of Random Samples of a Population," *Biometrika*, 1906, V. pp 172—5 (an exceedingly clear summary of the principles involved). Also later papers in *Phil Trans* 1901, CXCVII. A, pp. 443—459, and 1916, being supplements to the first mentioned memoir See also pp lx to lxx in Pearson's *Tables for Statisticians and Biometricians*, Cambridge, 1914.

† Compare equation (5), Chap. II, p 34

(3) Putting $c_3 = c_4 = \dots = 0$, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{x + a}{c_0 + c_1 x + c_2 x^2},$$

which can be made to represent almost all the frequency distributions which may arise. It fits the hypergeometrical series, the successive terms of which, e g give the chances of getting $k, k-1, \dots 0$ black balls from a bag containing pn black and qn white balls when k balls are drawn*.

Here the contributory causes are *not* independent of one another

There is no advantage in employing equations which involve c_3 and higher constants, because their use necessitates the calculation of the 6th and higher "moments," and these have very high probable errors

Definition. The *nth moment coefficient* (μ_n') of any distribution about *any* ordinate is the sum of the products of the partial frequencies and

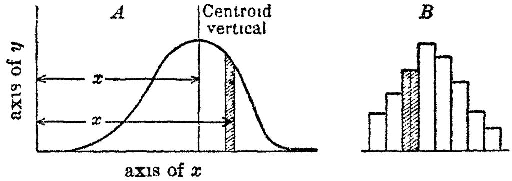


Fig 10

the nth power of the distances of these frequencies from the ordinate, divided by the total frequency. In symbols, if N be the total frequency,

$$N\mu_n' = \int x^n y \delta x.$$

The moments are, in practice, first calculated about any arbitrary ordinate that is most convenient, and then reduced to moments about the centroid in the following way (dashed μ's represent moments about an arbitrary ordinate, undashed μ's about the central ordinate):

$$N\mu_n = \int (x - \bar{x})^n y \delta x = N\mu_n' - n\bar{x}N\mu'_{n-1} + \frac{n(n-1)}{1.2}\bar{x}^2 N\mu'_{n-2} - \dots.$$

This gives the general reduction formula

$$\mu_n = \mu_n' - n\mu_1'\mu'_{n-1} + \frac{n(n-1)}{1.2}\mu_1'^2\mu'_{n-2} - \dots,$$

* The series is

$$\frac{pn(pn-1)\dots(pn-k+1)}{n(n-1)\dots(n-k+1)}\left\{1 + \frac{kqn}{pn-k+1} + \frac{k(k-1)}{2!}\cdot\frac{qn(qn-1)}{(pn-k+1)(pn-k+2)} + \dots\right\}$$

Other series may arise.

which enables us to transfer any moment from an arbitrary ordinate to the mean. Thus we have

$$\mu_2 = \mu_2' - \mu_1'^2,$$
$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3,$$
$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4.$$

The symbols μ represent moments of the *curve*, but we have to start with grouped frequencies, where the frequencies are assumed to be concentrated along the mid-ordinates of the rectangles (cf Fig. 10 $\hat{B}$). The moments obtained from these grouped frequencies are denoted by ν's (dashed and undashed), and corrections are necessary These have been deduced by Sheppard* and are consequently known as Sheppard's adjustments. They are

$$\mu_2 = \nu_2 - \tfrac{1}{12},$$
$$\mu_3 = \nu_3,$$
$$\mu_4 = \nu_4 - \tfrac{1}{2}\nu_2 + \tfrac{7}{240}.$$

It is generally said that they are only valid when there is "high contact" at the ends of the frequencies, but the equations for μ_2 and μ_3 are probably still valid even without high contact, if the terminal frequencies are zero.

(N B. In working, ν''s are changed into ν's *before* applying Sheppard's corrections.)

It is obvious that $\nu_0 = 1$ and $\nu_1 = 0$. ν_1' is the distance of the mean or centroid vertical from the arbitrary ordinate about which the moments are first taken, and is conveniently known as d.

Two very important constants in curve fitting are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

The values of these are always to be calculated, and, within the limits of their probable errors, they fix the type to which the curve belongs The general frequency curve equation, written in terms of moments, is

$$\frac{1}{y}\frac{dy}{dx} = -\frac{x + \sigma\dfrac{\sqrt{\beta_1}}{2}\dfrac{(\beta_2+3)}{5\beta_2 - 6\beta_1 - 9}}{\sigma^2\left\{\dfrac{4\beta_2 - 3\beta_1}{10\beta_2 - 12\beta_1 - 18} + \dfrac{\sqrt{\beta_1}}{2}\dfrac{(\beta_2+3)}{5\beta_2 - 6\beta_1 - 9}\cdot\dfrac{x}{\sigma} + \dfrac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}\left(\dfrac{x}{\sigma}\right)^2\right\}}.$$

* W F. Sheppard, "On the Calculation of the most Probable Values of Frequency Constants, for Data arranged according to Equidistant Divisions of a Scale," *Proc Lond. Math Soc.* xxix pp 353 ff Karl Pearson, "On an Elementary Proof of Sheppard's Formulae for correcting Raw Moments and on other allied Points," *Biometrika*, 1904, III. pp. 308 ff

This gives at once, for the distance between the mean and the mode,

$$x = -\frac{\sigma\sqrt{\beta_1}\,(\beta_2+3)}{2\,(5\beta_2-6\beta_1-9)}$$

(origin is at mean).

Hence the curve is symmetrical if $\beta_1 = 0$ If $\beta_1 = 0$ and $\beta_2 = 3$, the curve reduces to the Gaussian or Normal Curve, since the terms involving x in the denominator of the right-hand side of the general frequency curve equation then vanish

In using the Pearsonian method, then, the order of procedure to be adopted is.

(1) Calculate the moment coefficients ν_1', ν_2', ν_3', ν_4' about a convenient arbitrary ordinate,

(2) Transfer to the mean by the equations

$$\begin{aligned}\nu_2 &= \nu_2' - \nu_1'^2,\\ \nu_3 &= \nu_3' - 3\nu_1'\nu_2' + 2\nu_1'^3,\\ \nu_4 &= \nu_4' - 4\nu_1'\nu_3' + 6\nu_1'^2\nu_2' - 3\nu_1'^4.\end{aligned}$$

(ν_1' or d is the distance of the mean from the arbitrary ordinate)

(3) Determine the corresponding moments for the curve by the equations

$$\left.\begin{aligned}\mu_2 &= \nu_2 - \tfrac{1}{12}\\ \mu_3 &= \nu_3\\ \mu_4 &= \nu_4 - \tfrac{1}{2}\nu_2 + \tfrac{7}{240}\end{aligned}\right\}\text{Sheppard's corrections.}$$

(N B. For these corrections to be applicable, two conditions must be fulfilled.

(i) there must be high contact,

(ii) the grouping of the frequencies must be *equal.*)

(4) Calculate β_1 and β_2 by the equations

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

These results give the distance of the *mean* from the arbitrary ordinate (ν_1' or d), the *standard deviation* ($\sqrt{\mu_2}$), and the *mode*

$$\text{mean} - \frac{\sigma\sqrt{\beta_1}\,(\beta_2+3)}{2\,(5\beta_2-6\beta_1-9)}.$$

The *median* is more difficult to determine exactly, but a position which is approximate and indeed very accurate in all the curves we are likely to meet with in this work, is between the mean and the mode, but

nearer to the mean, so that the distance from the mode is twice the distance from the mean*

$$\text{Mean}\underbrace{\qquad\quad}_{1/3}\text{Median}\underbrace{\qquad\quad}_{2/3}\text{Mode}$$

The investigator may then proceed to determine to which of the Pearsonian "types" the particular curve belongs, to find its equation, and to plot it. The type is decided by the constants β_1 and β_2 (using Diagram XXXV, p. 66, Pearson's *Tables*) or by the criterion κ_2 where

$$\kappa_2 = \frac{\beta_1 (\beta_2 + 3)^2}{4 (4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)},$$

using the following diagram:

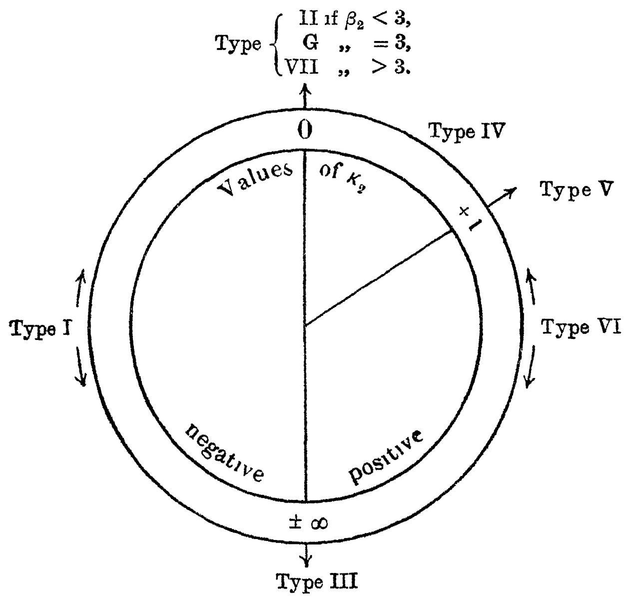


Fig 11

The equations to the types are given in detail in Pearson's *Tables*, p. lxiii, or in Elderton (*op. cit.*), where valuable advice on the arrangement of the calculations will be found. Elderton's Type II includes

* Karl Pearson, *Biometrika*, I 1902—3, p. 265, *Phil. Trans* CLXXXVI. A, p. 375; *Roy. Soc. Proc* LXVIII p 369 b C V. L Charlier, "Researches into the Theory of Probability," *Lunds Universitets Årsskrift*, 1905 (1), 1, equation 9 Arthur T. Doodson, "Relation of the Mode, Median, and Mean in Frequency Curves," *Biometrika*, 1917, XI p. 425

both our Types II and VII, and his Type VII is our G, the Normal Curve

We shall confine ourselves to a fuller description of Type IV, which appears to be the type most common in psychometric skewness*.

It will be clear from the above account that the whole of the calculations are based upon the first four *moments* of the data, and we proceed first to describe the most convenient way of finding these when as here the stimuli are equidistant, viz the summation method already used for finding the mean on p. 19 If the stimuli are not equidistant the calculations are rather longer.

(5) THE SUMMATION METHOD OF FINDING MOMENTS IN THE CASE OF DATA AT EQUIDISTANT POINTS

A full description of this device will be found in Mr Palin Elderton's book already cited, where it is attributed to Mr G F. Hardy. It has however been independently used by numerous writers, e g. Lipps, Wirth, Urban, etc.†

The theory cannot be worked out here, but the reader can easily do so for himself It is only a question of simple algebra; or from another point of view it is the same thing as integration by parts. We give only a worked example, Urban's Subject I, *heavier* answers.

Example of the Use of the Summation Method

Grams	p	series of successive sums		
84	0022	0022	0022	0022
88	0200	0222	0244	0266
92	0889	1111	1355	·1621
96	2222	3333	4688	6309
100	·4133	7466	1 2154	1 8463
104	8956	1 6422	2 8576	4 7039
108	9400	2 5822	5 4398	10 1437
	2 5822	5 4398	10 1437	17 5157
	S_2 or d	S_3	S_4	S_5

* Thirteen out of fifteen psychometric curves tried recently were Type IV Cf Pearson on zoological and anthropological curves, where Type IV also prevails, *Phil Trans* 1895, CLXXXVI A, Part I, pp 388, 403 and 411

† G. F Lipps, "Die Theorie der Collectivgegenstande," Wundt's *Phil Stud* Bd XVII Separat-Abdruck, Leipzig, 1902, W Wirth, "Die mathematischen Grundlagen der sogenannten unmittelbaren Behandlung psychophysischer Resultate," Wundt's *Psychol. Studien*, 1910, Bd. VI. pp 141, 252, 430 Urban on Wirth, *Archiv f. d. ges Psychol.* XX Literaturbericht, p. 1.

The table is self-explanatory. Each column is formed from the preceding one by successive summations (from the top in this case), and is then totalled*. The origin is here the centre of the group beyond 108 grams, viz 110 grams, and the unit of measurement is 4 grams, measured downwards from 110 grams. We have

$$\text{Mean} = 110 - 4d = 99{\cdot}6712 \text{ grams.}$$

Further, it can be shown that the moments

$$\begin{aligned}
\nu_2 &= 2S_3 - d(1+d),\\
\nu_3 &= 6S_4 - 3\nu_2(1+d) - d(1+d)(2+d),\\
\nu_4 &= 24S_5 - 2\nu_3\{2(1+d)+1\}\\
&\qquad - \nu_2\{6(1+d)(2+d)-1\} - d(1+d)(2+d)(3+d).
\end{aligned}$$

The work is not heavy up to this point if arranged systematically, and in the present case it gives

$$\begin{aligned}
\nu_2 &= 1{\cdot}6296,\\
\nu_3 &= 0\ 9645,\\
\nu_4 &= 9{\cdot}1621,
\end{aligned}$$

or, using Sheppard's corrections,

$$\begin{aligned}
\mu_2 &= 1{\cdot}5463, \qquad \sigma = \sqrt{\mu_2} = 1{\cdot}2435,\\
\mu_3 &= 0{\cdot}9645,\\
\mu_4 &= 8{\cdot}3765.
\end{aligned}$$

This σ gives in original units $4 \times 1{\cdot}2435 = 4{\cdot}974$ grams, differing from the value mentioned on p. 63 because of the Sheppard adjustment. From the moments we obtain

$$\begin{aligned}
\beta_1 &= 0{\cdot}251,\\
\beta_2 &= 3{\cdot}51,\\
k_2 &= 0{\cdot}775.
\end{aligned}$$

The type is therefore Type IV; within the limits of probable error of β_1, β_2, and κ_2 it might however be Type G, VII, or V. Unfortunately the ordinary methods of finding these probable errors are of doubtful significance in the case of pseudo-histogram data such as ours. We turn to the calculation of Type IV.

* Slightly greater accuracy could, by the way, be attained by using the actual numbers of answers *heavier* and not the proportions p in cases like the present where an awkward number of experiments was performed at each stimulus, viz. 450, leading to recurring decimals The totals would then be divided by this number

(6) CALCULATION OF A SKEW CURVE (Type IV)

The equation* is

$$y = y_0\left(1+\frac{x^2}{a^2}\right)^{-m} e^{-\nu\tan^{-1}\frac{x}{a}},$$

wherein $m = \frac{1}{2}(r+2)$ say,

where $r = 6(\beta_2-\beta_1-1)/(2\beta_2-3\beta_1-6)$,

$$\nu = \frac{r(r-2)\sqrt{\beta_1}}{\sqrt{\{16(r-1)-\beta_1(r-2)^2\}}} = rl/n \text{ say},$$

$$a = n\sqrt{\mu_2}/4,$$

Sk (skewness) $= l/(4m)$.

Origin, at mean $+ \nu a/r$;

mode, at mean $-$ skewness $\times\, \sigma$;

$$y_0 = \frac{N}{a}\frac{\sqrt{r}}{\sqrt{(2\pi)}}\frac{e^{\frac{\cos^2\phi}{3r}-\frac{1}{12r}-\phi\nu}}{(\cos\phi)^{r+1}},\quad \tan\phi = \frac{\nu}{r}.$$

The actual form of the calculation depends on the appliances available, whether Crelle's Tables, or logarithms, or calculating machines Using Crelle and seven figure logarithms we get finally, after much labour, the following values for the ogive, which has to be obtained from the bell-curve by simple quadrature:

Urban's Subject I, Heavier *answers*

Stimulus	Observed p	Calculated by Type IV	$(p-p')^2/(p'q')$
84	0022	0046	0012
88	0200	0184	0001
92	0889	0690	0062
96	2222	2155	0002
100	4133	5006	0305
104	8956	·8078	0496
108	·9400	9624	0139
			Sum 1017

$$\chi^2 = 450 \times \cdot 1017 = 45\ 8,\quad P < \ 0000005 \text{ still.}$$

We find therefore that fitting the best Pearson curve possible to the data makes practically no improvement in the fit, which is still so very bad as to make it quite certain that the data are not homogeneous at all. The fit cannot be much improved by other assumptions as to the spread of the "tail," several of which have been tried The Type IV

* The notation on this page is that of Elderton following Pearson, and the symbols m, ν, n and r have no connection with these symbols used elsewhere in the present book.

curve itself is shown, contrasted with the pseudo-histogram, in Fig. 12. The mean (99·67), median (99·97) and mode (100·52) are worth comparing, as a matter of interest, with the thresholds obtained otherwise (see pp. 53, 60, 63 and 70)

The reason for the bad fit in this case is, mathematically, the impossibility of finding a curve to accommodate both the tall rectangle 217, and the tail of 27, however the latter may be allocated. Much of Urban's other data shows the same bad fit, and for the same reason, the size of the "tails" Not all however are bad The best case is Subject II

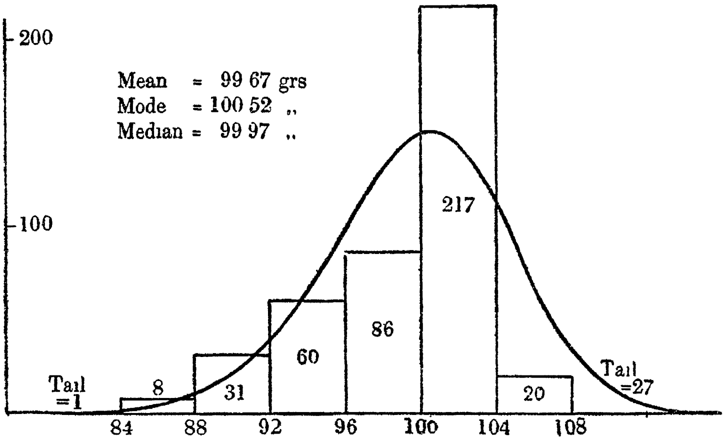


Fig 12. A Type IV curve fitted to Urban's data for Subject I, *heavier* answers, in pseudo-histogram form

(Urban himself) *who had had much practice at this form of experiment.* The *lighter* answers in his case were as follows, compared (1) with a curve fitted by the Constant Process and (2) with a Pearson skew curve (here Type I).

Urban's Subject II, Lighter *answers*

Grams	Observed p	Normal Curve p	Type I Curve p
84	·9333	·9504	·9432
88	·8622	·8540	·8520
92	·7000	·6767	·6875
96	·4489	4456	·4627
100	·2311	2320	·2379
104	0956	·0922	·0858
108	·0156	0272	·0187

On testing the goodness of fit we obtain

For Normal Curve, $P = 0{\cdot}48$.

For Type I Curve, $P = 0{\cdot}91$.

Here the Gaussian is a good, and Type I an excellent fit. There is every reason then to think that the data here are homogeneous The bell-curve and pseudo-histogram are shown in Fig 13

Since we have decided that the data of Urban's Subject I are heterogeneous, that is, that the conditions of the experiment varied considerably during its performance (which lasted several months), the question naturally arises as to whether we can analyse the data mathematically into two or more frequency-distributions. This question was discussed by Professor Pearson in 1894* as far as an analysis into two normal curves goes One more moment, μ_5, is needed and as the probable error of this is considerable the practical application of the plan to be described is not very satisfactory.

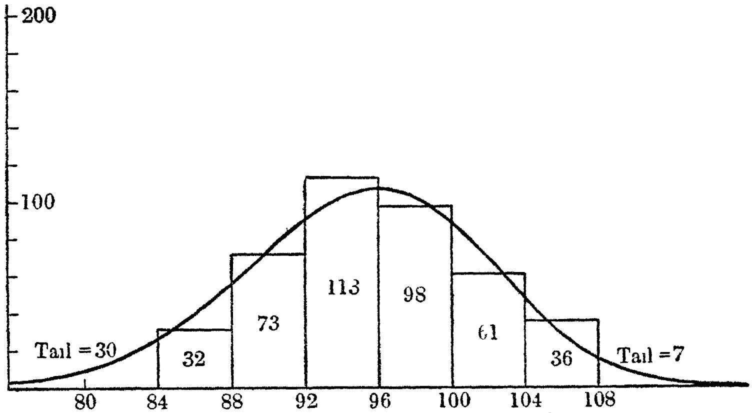


Fig. 13. A Type I curve Urban's Subject II, lighter answers

(7) ANALYSIS INTO TWO NORMAL CURVES

Stage I. Find the centroid of the frequency curve and calculate $\mu_2, \mu_3, \mu_4, \mu_5, \lambda_4$ and λ_5.

$$\lambda_4 = 9\mu_2^2 - 3\mu_4,$$
$$\lambda_5 = 30\mu_2\mu_3 - 3\mu_5.$$

Stage II. Solve the following nonic equation for p_2 using Sturm's functions to localise roots:

$$24p_2^9 - 28\lambda_4 p_2^7 + 36\mu_3^2 p_2^6 - (24\mu_3\lambda_5 - 10\lambda_4^2)\, p_2^5 - (148\mu_3^2\lambda_4 + 2\lambda_5^2)\, p_2^4 + (288\mu_3^4 - 12\lambda_4\lambda_5\mu_3 - \lambda_4^3)\, p_2^3 + (24\mu_3^2\lambda_5 - 7\mu_3^2\lambda_4^2)\, p_2^2 + 32\mu_3^4\lambda_4 p_2 - 24\mu_3^6 = 0,$$

and find p_1 from

$$p_1p_2 = \frac{2\mu_2^3 - 2\mu_3\lambda_4 p_2 - \lambda_5 p_2^2 - 8\mu_3 p_2^3}{4\mu_3^2 - \lambda_4 p_2 + 2p_2^3}$$

* *Phil Trans. Roy Soc. London*, 1894

Stage III. Find γ_1 and γ_2 the roots of

$$\gamma^2 - p_1\gamma + p_2 = 0.$$

$h\gamma_1$ and $h\gamma_2$ are the positions of the axes of the normal component curves, where h is the unit of length.

Stage IV. The fractions z_1 and z_2, that the areas of the component curves are of the area of the whole curve, form the roots of the quadratic

$$z^2 - z - \frac{p_2}{p_1^2 - 4p_2} = 0.$$

Stage V. The standard deviations are found from

$$\sigma_1^2/h^2 = \mu_2 - \tfrac{1}{3}\mu_3/\gamma_2 - \tfrac{1}{3}p_1\gamma_1 + p_2,$$

$$\sigma_2^2/h^2 = \mu_2 - \tfrac{1}{3}\mu_3/\gamma_1 - \tfrac{1}{3}p_1\gamma_2 + p_2.$$

In the case of our curves, the fact that the tails are only known as to area and not as to distribution makes this procedure hardly worth while, for the value of μ_5, which in any case has a high probable error, depends here to a very great extent on how these tails are allocated Trials however have shown that the decrease in the value of χ^2 obtained by dissecting into two normal curves is only very slight indeed The heterogeneity is more complex than can be thus dealt with

Note, 1924. In the *Am Journ Psychol* 1923, F M Urban suggests, and S W Fernberger carries out, calculations on some of the latter's data for the purpose of testing the suggestion made on p 90 above, by Hoisington in *Am Journ Psychol* 1917 (and, Urban might have added, by himself much earlier), that practice brings about an approximation to the normal curve of distribution, or $\phi(\gamma)$ hypothesis as American writers call it The calculations were in favour of this being the case In the same *Journal* in 1920 E G Boring published an important article on the logic of the normal law in mental measurement, to which an equally important reply by T L Kelley appeared in 1923 Boring, who rightly denies any inherent virtue in the normal distribution, is particularly concerned about the impossibility of everything being scattered in this way As he says, the cubical crystals of common salt cannot be normally distributed both as regards height and as regards weight, since the latter measure is proportional to the cube of the former Attempts to force mental measurements to fit a normal curve and thereby to deduce a system of units find Boring sceptical Such attempts were made by Galton and by Pearson and more recently by Trabue and others in America where McCall with his "T-scale" has given the device wide popularity.

As a consequence of his rejection of the normal curve Boring rejects all psychological units except "sense-distances." We are left then, he says, with rank-orders, medians, quartiles, contingencies and correlation-ratios instead of measurements, averages, standard deviations, coefficients of correlation, and linear regression Kelley however retorts that

---

Figures 14, 22 and 23 which appeared here and on pages 124, 128 in the previous edition have been omitted, but remaining figures retain their original numbers.

(8) CONCLUSIONS

Further analysis is useless For since there are only seven points given by experiment in the curve, and also the total area, it is clear that an *exact* fit could be obtained by two skew curves, or a skew and a normal, which have between them eight constants, or in many ways by three normal curves

All we can say is that this subject's sensitivity oscillated between *at least* two states, and if only two, then one of these is such as to give a skew curve. Without attempting to get an exact fit, it is clear after our experience, that provided we supply a flat normal curve to give the awkward tails, the surplus of the distribution could be fitted successfully by a skew curve

This suggests that the two components into which the heterogeneous data are thus divided are (*a*) a component due to erratic answers giving a wide shallow distribution and (*b*) a component really corresponding to the conditions of the experiment

The use of "catch" tests in threshold determination is to check component (*a*). The proper way to employ them is to cancel sittings at which numerous catch errors occur, the subject being then presumably in an abnormal state.

In Riecker's experiment many "catch errors" were present, i e

a correlation ratio involves standard deviations, so it too must go, if Boring is to be consistent and by less obvious but as I think valid arguments shows that none of the other measures retained by Boring is entirely free of those things that he inveighs against In the second part of his paper Kelley considers in more detail what are the defensible bases for determining the units of a mental scale He anticipates that much greater difficulty will be experienced in determining a homogeneous ordered series than in scaling it afterward, and lets fall the interesting hint that he hopes at a later date to offer criteria for determining if a number of mental tasks involve one or more qualitatively different mental functions, a problem identical with or at least similar to that discussed in Chapters IX and X of this book

In the second part of his article Kelley suggests four units of value in mental measurement (*a*) the sensed difference unit, (*b*) the variability in performance unit, (*c*) the group variability unit, and (*d*) the unit resulting in the simplest picture of interrelationships, as e g stretching and compressing units till regressions are linear

Very interesting is Kelley's method of arriving at a psychometric function by successive approximations If the first guess is right it will lead to the same measure of $\overline{ab}$ whether the standard stimulus be a, b, c or d The wrong guess will give different values of $\overline{ab}$ which Kelley averages and he uses this, and the averages of $\overline{ac}$, $\overline{ad}$, etc , as points on a better psychometric function from which he again goes through the same procedure and so on.

occasions when the subject answered *two* to a one-point stimulus Fitting the best normal curve to his data, we only get, on testing its goodness of fit, $P = \cdot 0003$

Riecker's Data

Mean 2 02, $\sigma = 1\ 46$ Paris Lines

Paris Lines s	Observed p	$s-2\ 02$	$\frac{s-2\ 02}{1\ 46}$	Gaussian p'	$p-p'$	$(p-p')^2/p'q'$
0	00	−2 02	−1 38	084	− 084	092
0 5	·10	−1 52	−1 04	149	− 049	·019
1	·14	−1 02	−0 70	242	− 102	057
1 5	·40	−0 52	−0 36	·359	041	·007
2	·65	−0 02	−0 014	·494	·156	097
3	80	0 98	0 67	·749	051	·014
4	87	1 98	1 36	913	− 043	023
5	96	2 98	2 04	979	− 019	018
6	1 00	3 98	2 73	997	003	003
					Sum	·330

$\chi^2 = 100 \times 330,\ n' = 10,\ P = 00031$

Compare with this the following case The data are for the spatial threshold on the forearm*, and were gathered by the *Method of Non-Consecutive Groups* Sittings containing more than a certain number of catch errors were rejected, the number chosen being one which when exceeded at all was usually exceeded violently. Other experimental precautions were taken which are described in the articles cited. As a result it is found that a roughly fitted Normal Curve is quite a fair fit, for in 13 out of 100 cases a worse departure would be got by chance: and a skew curve would improve this. The data are in fact reasonably homogeneous. (Calculations on opposite page.)

The conclusion of the whole matter is that we are led to believe that the difficulties of psychophysical experiment are such that homogeneity in the data is rare. For such data refinements of mathematical calculation are out of place. The curve fitting methods here described are however of value in *discovering* the heterogeneity†.

With increasing precautions in carrying out the experiments and with increasing practice on the part of the subject, it would appear that the data finally reach a distribution where they are fairly well fitted by a Normal Curve, and excellently fitted by Pearsonian Skew Curves

* G. H. Thomson "A Comparison of Psychophysical Methods," *Brit Journ. Psychol* 1912, v. pp 203—241; "Changes in the Spatial Threshold at a Sitting," *Brit. Journ. Psychol* 1914, vi. pp. 432—448 and *B.A Report*, 1913, pp. 681—683

† Compare Urban's use of the Lexian Coefficient of Dispersion (op cit.) and compare the latter with Pearson's more significant criteria β_1, β_2, and the $\beta_1\beta_2$ diagram (XXXV in his *Tables*)

Normal Integral fitted to Thomson's Spatial Threshold Data

Cms	Two-point answers	Continued sum	Cms from mean	In σ units	Theoretical p' from Sheppard	Observed p	$(p-p')^2/p'q'$
0	5 2	5 2	−2 44	−2 29	011	035	05293
$\frac{1}{2}$	6	11 2	−1 94	−1 82	034	040	00109
1	8	19 2	−1 44	−1 35	089	053	01599
$1\frac{1}{2}$	21	40 2	−0 94	−0 88	189	140	·01565
2	56	96 2	−0 44	−0 41	341	374	00485
$2\frac{1}{2}$	84	180 2	0 06	0 06	524	560	00521
3	105	285 2	0 56	0 52	698	700	00002
$3\frac{1}{2}$	125	410 2	1 06	0 99	839	834	00018
4	141	551 2	1 56	1 46	928	940	00216
$4\frac{1}{2}$	144	695 2	2 06	1 93	973	960	00644
5	148	843 2	2 56	2 40	992	957	00315
150 { 5	843 2	3137 2				Sum	=·10767
150 { 30	168 64	627 44					
	5 621 d	20 915 S_3					

d in cms is 2 81 from an origin of $5\frac{1}{4}$ downwards

Mean $=5\ 25-2\ 81=\ 2\ 44$ cms

$$2S_3=41\ 83$$
$$d\,(1+d)=37\ 20$$
$$\nu_2=\ 4\ 63$$
Sheppard's correction $\ 0\ 08$
$$\mu_2=\ 4\ 55=\sigma^2$$
$$\sigma=2\ 133$$
or 1 067 cms.

$$\chi^2=150\times\ 10767=16\ 15,$$
$$n'=12,\ \ P=0\ 13$$

PART II

CORRELATION

CHAPTER V

INTRODUCTION TO CORRELATION

A SOMEWHAT detailed account of the mathematical theory of correlation and of the way in which it may be usefully applied to psychological measurements will be found in the later chapters of this Part The object of the following introductory pages is to give the reader a general preliminary view of the method, free from mathematical complications, and to illustrate it by means of a simple example.

Correlation may be briefly defined as "tendency towards concomitant variation," and a so-called correlation coefficient (or, again, correlation ratio) is simply a measure of such tendency, more or less adequate according to the circumstances of the case J. S Mill, in his "System of Logic," distinguished a special scientific "Method of Concomitant Variations," which he based upon the following principle

"Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation*."

The instances of this principle which Mill had in mind were mainly cases of approximately "complete" concomitance of variation, such as those usually met with in the domain of Physics In such cases, the conditions of an experiment admit of a high degree of simplification, the phenomenon, or series of phenomena, under investigation can be isolated with tolerably complete success, and the "irrelevant" factors can be reduced to a minimum Under such conditions, when the degree of concomitance of the different corresponding measures of the two phenomena is found to be very high, the slight deviations from complete

* *Logic*, Bk. III. Ch. VIII § 6.

correspondence are put down to "errors of observation" or other unavoidable imperfections in the experimental method employed

If the correspondence is one of simple proportionality, so that the graphical representation of it (one phenomenon being measured along the axis of x, the other along the axis of y) is a straight line*, the correlation coefficient r will be unity *Example* the variation of the length of a metal rod with temperature

If the correspondence although still approximately complete, is not one of simple proportionality, the graphical representation of it will be, not a straight line, but a curve of greater or less complexity†, and the correlation, also complete, will be measured not by the correlation

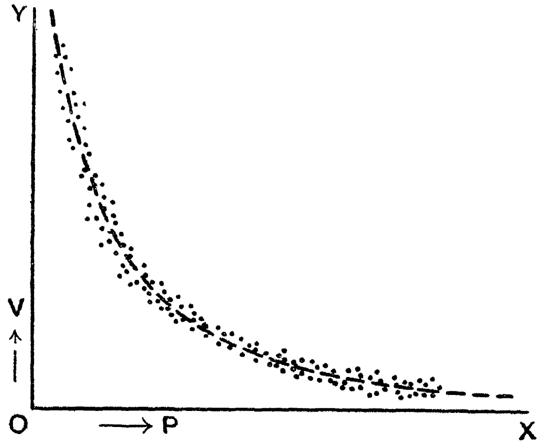


Fig 15

coefficient r, but by the correlation *ratio* η, which in this case will be unity. *Example*: the variation of the volume of a certain quantity of gas with the pressure to which it is subjected, the temperature remaining constant A number of pairs of values P_1, V_1, P_2, V_2; P_3, V_3, etc. is obtained, and when plotted they are found to give a "scatter diagram" of the approximate form of Fig. 15.

In this figure, the dots represent the individual pairs of observations P, V. They cluster very closely about the hyperbola, $PV = k$, represented by the broken curve. The curve is assumed to represent the "real" or "true" relation of the two "variates" (as we call such quantities as P, V), and the slight deviations of the observed values

* Hence the correlation is said to be "linear"

† Correlation said to be "non-linear" or "skew."

from this curve are explained as due to errors of observation and to other factors irrelevant to the relation under investigation However this may be, the interesting point about the figure so far as our present purpose of explaining correlation is concerned is that any definite observed P-value is "correlated" with a plurality or "array" of observed V-values, and that, similarly, any definite observed V-value is correlated with a plurality of P-values These arrays of observed values cluster extremely closely about their means (situated on the curve), i e their "scatter" or "variability," as measured by their standard deviations (σ), is extremely small

The modern theory of correlation is directed towards the manipulation of observations made upon phenomena of a much greater degree of variability than that found in the case of isolated physical phenomena. The increased variability is no doubt due, in the main, to the complexity of factors involved The elementary factors do not admit of isolation, and with reference to the concomitance of variation of the two series of phenomena under consideration they, as it were, pull in different directions The correlation coefficient and correlation ratio measure, in these cases, the average extent of the concomitance. As will be explained more fully in the next chapter, r can only be taken as a measure of correlation when the average relation between the two variates is *linear*, and in this case its value is identical with that of η When the relation is non-linear r is practically meaningless, but η still measures the relation accurately.

The general problem will become clearer by reference to the accompanying figure.

Let us assume that we have a group of 200 school-children and have measured each of them for mechanical memory (x) and for general intelligence (y) Each of the dots in the figure represents a child Then if we determine the mean y-values corresponding to each successive "group" of x-values, e g. x_2 to x_3, by assuming the observations concentrated on the mid-ordinate PM*, the line AB† drawn through these mean y-values (marked by crosses in heavy type) represents the law of change of mean y-value with increase of x and gives the "most probable"

* The true centroid ordinate is slightly nearer the denser part of the scatter diagram, here slightly towards the right of PM. The correction is made later by means of Sheppard's formulae (see above, p 84)

† The means have been placed on the straight line for the sake of convenience of exposition. Actually, they will occur irregularly on either side of it, and AB will be the "best fitting" straight line, determined by an application of the Method of Least Squares See next chapter.

value of y for any particular value of x If the line is straight or approximately straight the "regression" is said to be linear, and the equation to the line is

$$y - \bar{y} = r\frac{\sigma_2}{\sigma_1}(x - \bar{x}),$$

where $\bar{x}$, $\bar{y}$ are the mean values of all the x's and y's respectively (*not* the means of the arrays just mentioned), σ_1, σ_2 are the standard deviations of the x's and y's respectively, and r is the coefficient of correlation.

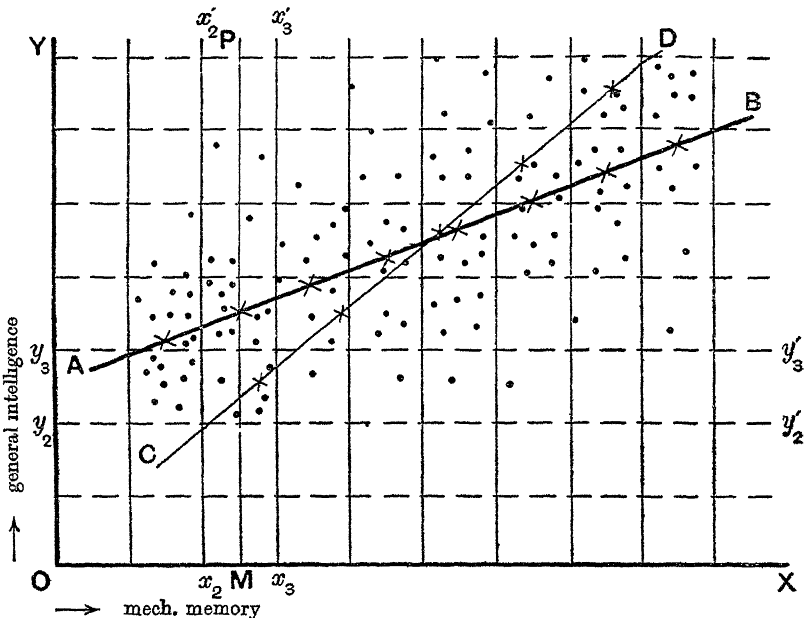


Fig 16

Thus if

20 is the mean value of the mechanical memory of the group,
30 ,, ,, ,, ,, general intelligence of the group,
4 ,, ,, standard deviation for mechanical memory (σ_1),
7 ,, ,, ,, ,, ,, general intelligence (σ_2),

and, finally, ·6 is the value of r; then the most probable measure of the general intelligence of a child whose mechanical memory is represented by, say, the value 14, is given by the equation

$$y - 30 = \cdot 6 \times \tfrac{7}{4}(14 - 20),$$

whence $y = \mathbf{23{\cdot}7}$. This value is the average of an array of possible values, whose standard deviation

$$= \sigma_2 \sqrt{(1 - r^2)}$$
$$= 5\ 6.$$

It will be proved in the next chapter that

$$r = \frac{S\,(xy)}{N\sigma_1\sigma_2}*,$$

where x and y are deviations from the mean (not absolute values as assumed above), and S ( ) indicates summation, i e

$$S\,(xy) = x_1y_1 + x_2y_2 + x_3y_3 + \quad + x_Ny_N,$$

where N is the total number of cases (children measured)

It is important to note that by starting from y instead of from x, and determining the means of the y-arrays (such as the array within the limits y_2y_2', y_3y_3'), another regression line, CD, is obtained *different* from the first Its equation is

$$x - \bar{x} = r\,\frac{\sigma_1}{\sigma_2}\,(y - \bar{y}),$$

and it represents the law of change of mean x-value with increase of y It gives the "most probable" value of x for any particular value of y

If the series of means do not lie on a straight line (approx) but on a curve of greater or less complexity, the above calculation is meaningless. In such a case, called a case of skew correlation and non-linear regression, the only measure of the correlation of the two variates is that given by η, the correlation ratio η is the ratio of the standard deviation of the means of the arrays (Σ) to the total standard deviation (of either the x's or the y's). Thus there are two values of η, one for the x's, and another for the y's They approximate closely to one another, as a rule, so that only one need be calculated.

$$\eta = \frac{\Sigma_1}{\sigma_1}, \text{ or } \frac{\Sigma_2}{\sigma_2}.$$

When the regression is linear, $\eta = r$, otherwise $\eta > r$ r ranges between the values ± 1, η between 0 and 1. η is always positive

It will now have become clear that the correlation ratio, η (always)

* First suggested by Bravais, shown to be the best measure by Professor Karl Pearson, who gave it the name of the "product moment" formula A. Bravais, "Analyse mathématique sur les probabilités des erreurs de situation d'un point," *Acad des Sciences, Mémoires présentés par divers savants*, IIe Série, IX 1846, p 255 Karl Pearson, F R S, "Regression, Heredity and Panmixia," *Phil. Trans Roy Soc*, 1896, CLXXXVII A, pp 253 ff. But see Pearson, *Biometrika*, 1920, XIII. p 25.

and the correlation coefficient, r (when regression is linear) are measures of the tendency towards concomitant variation exhibited by two series of phenomena, and hence throw some light upon the causal relations of these phenomena. Exactly what kind of causal relation we are justified in inferring from them will become clearer in the course of the next few chapters

We may illustrate the significance of the idea of correlation in a slightly different (and more elementary) way. Let us suppose that the 200 children have been arranged in order of merit, as regards mechanical memory, on the one hand, and as regards general intelligence on the other If now it were found that each child's order was the same in both, i e that the child first in mechanical memory was first in general intelligence, the child second in mechanical memory was second in general intelligence, and so on, correspondence between the two series would be complete and r would equal $+ 1$ Or if, on a second supposition, the child first in the one was last in the other, the child second in the one was next to last in the other, and so on, the correspondence between the two series would again be complete, but inverse, and r would be $- 1$. Finally, if there is no correspondence whatever between the two series, r will be zero. A value of r between 0 and $+ 1$ will express a tendency, greater or less according to r's size, for children above the average or mean position in the one ability to be above the mean position in the other, and for children below the mean position in the one to be below the mean position in the other. A value of r between 0 and $- 1$ will express a tendency, greater or less according as r is numerically greater or less, for the children above the mean position in the one ability to be below the mean position in the other, and conversely. Now if order or "rank" be taken as an inverse measure of ability, the value of

$$\frac{S\,(xy)}{N\sigma_1\sigma_2} \text{ or } r$$

becomes

$$1 - \frac{6S\,(d^2)}{N\,(N^2 - 1)},$$

where d is the difference between the rank of an individual in the one series and his rank in the other. This form gives us a general impression of its appropriateness for the purpose in view, since the greater the disparity between the two series of ranks the greater is $S\,(d^2)$ and hence the smaller is r. If there is no relation at all between the two series, $S\,(d^2)$ acquires the value it would have according to pure chance, and this can be shown to be $N\,(N^2 - 1)/6$, which makes the whole expression zero, as it should do.

The one objection to the formula is that it assumes the difference between any two neighbouring ranks to be equal at all parts of the scale This is obviously a false assumption, the distance of individual from individual at the two extreme ends of the scale must be considerably greater than that between individuals near the middle A correction for this, based on the assumption that the form of distribution of the abilities in each of the cases is Normal, has been calculated by Professor Pearson. It is

$$r = 2 \sin \left(\frac{\pi}{6} \rho\right),$$

where

$$\rho = 1 - \frac{6S(d^2)}{N(N^2 - 1)}.$$

At the end of this chapter is given a table whereby ρ-values may at once be converted into corresponding r-values, according to the above equation.

Finally, there is the question of the "probable error" (P E). Like every other constant calculated from a limited sample of variable material, the coefficient of correlation varies in value from sample to sample, and a measure is needed of the limits within which it may be expected with a fair degree of probability to lie This measure is given by the probable error. In the case of r determined by the product-moment formula, when N is sufficiently large,

$$\text{P E} = \frac{\cdot 67449}{\sqrt{N}}(1 - r^2),$$

which means that it is an even chance that the true value of r lies between the limits

$$r \pm \frac{\cdot 67449 (1 - r^2)}{\sqrt{N}}.$$

The chances are 16 to 1 against the value falling outside the limits $r \pm 3$ P.E.

For r determined by the rank formula, the probable error is slightly larger, being $\cdot 7063 (1 - r^2)/\sqrt{N}$.

If N, the number of cases, be small (say, less than 30), the probable error is larger. Its exact size under such conditions is not known

The following is an example of the way in which a correlation coefficient may be obtained by means of ranks. The subjects were boys in the Fourth Form of a Public School, and the correlation to be obtained is that between ability in Classics and ability in Drawing.

	Form Order: Classics	Form Order: Drawing	d^2
R C. O.	1	9	$(1 \sim 9)^2 =$ 64
H G M	2	2	0
B L	9	16	49
F L S	7	6	1
C. M S.	3	15	144
C J L H.	5	4	1
A L. P.	6	17	121
E G T.	4	3	1
F C F.	8	5	9
N P R. N.	11	14	9
H B D	10	12	4
S H T	14	7	49
H B. M	12	1	121
L H S	13	8	25
J P C.	15	10	25
E W.	16	18	4
C C M	17	11	36
L. H W.	18	13	25
E. M. J.	19	19	0
$N=19$			$688 = S(d^2)$

$$\rho = 1 - \frac{6S(d^2)}{N(N^2-1)} = 1 - \frac{6 \times 688}{19 \times 360} = \cdot 40,$$

$$\therefore \ r = 2 \sin\left(\frac{\pi}{6}\rho\right) = \cdot 416,$$

$$\text{P.E.} = \frac{\cdot 7063\,(1-r^2)}{\sqrt{N}} = \cdot 134.$$

r is here just over three times its probable error, and we might therefore feel inclined to conclude that it proves a real correlation between the two series. We must remember, however, that 19 is a very small number of cases, and that therefore the real probable error is considerably larger than that given by the formula. Hence the reality of the correlation is not so certain. Our caution is proved to be justified when we turn to the next higher form, the Remove, and find that, with the same number of boys, the correlation between ability for Classics and Drawing ability works out as $-\cdot 313$ ($\pm \cdot 14$), quite a different result It might be objected that other factors than mere smallness in the number of cases were responsible for the difference, e g. that the tendency to specialise in Classics was greater in the Remove than in the Fourth, and that the consequent neglect of Drawing by the abler boys lowered the correlation. To this it may be replied, firstly, that the drawing-master was the same for both forms, and was likely to get as much out of the boys as possible in each case, and, secondly, that the difference between

the two forms in respect of the degree of specialising tendency was insufficient to account for the disparity of the results.

The correct way to compare the results mathematically is to determine the *probable error of their difference*. This = the square root of the sum of the squares of the probable errors of each*, i.e.

$$\text{P.E}_{a-b} = \sqrt{\text{P E}_a^2 + \text{P E}_b^2},$$

which, in this case,

$$= \sqrt{\cdot 13^2 + \cdot 14^2}$$
$$= \cdot 19.$$

The difference $= \cdot 416 + \cdot 313 = \cdot 73$, nearly four times the size of its probable error $= \cdot 19$.

A very important extension of the theory of correlation is the conception of "partial" correlation. If, e g, three mental abilities are correlated with one another, it is of interest to know how closely any two of them are correlated with one another *for a constant value of the third*. Such a coefficient is written, in Yule's notation, $r_{12.3}$

This may be illustrated from our example by taking the form order for English into consideration in addition to that for Classics and that for Drawing The correlation between Classics and English works out as ·78, that between Drawing and English, as ·21.

Then the correlation between Classics and Drawing for "English constant" is

$$r_{CD\,E} = \frac{r_{CD} - r_{CE} r_{DE}}{\sqrt{(1 - r_{CE}^2)}\sqrt{(1 - r_{DE}^2)}}$$
$$= \frac{\cdot 42 - \cdot 78 \times \cdot 21}{\sqrt{(1 - \cdot 78^2)}\sqrt{(1 - \cdot 21^2)}} = \cdot \mathbf{42}.$$

Thus in this particular case, the "partial" coefficient is practically identical with the "entire" coefficient

If therefore boys were selected, out of a population of which the actual form is a random sample, so as to be all equal in their "English" ability, the correlation between their "Drawing" ability and their "Classics" ability would be unaffected. Of course such a set of boys, in addition to being all alike in English, would be less scattered in both Classics and Drawing (especially in the former) than are the boys of the actual form, and their average ability in these subjects would be higher or lower than that of the actual form according to the level of ability in English at which they had been selected

On the other hand the partial correlation of English and Drawing

* See p. 24.

for "constant Classics" will be found to be $-\cdot 2$, so that selection for Classics creates a negative correlation between English and Drawing, in so far as we can judge from this particular case.

The reader must be warned against the temptation to draw deductions as to the "common factors" uniting any of these pairs of subjects. The fallacy underlying such reasoning is discussed in pp. 139—145.

The principle of partial correlation can be extended to include an indefinite number of variables, and general formulae for this purpose will be given in Chapter VII.

It is obvious that when the subjects in the group examined are not all alike in respect of some irrelevant factor such as age, these same formulae can be employed to ascertain what the correlations would be in a group which was homogeneous with regard to the factor in question. Great care has to be taken in interpreting the results of such calculations however, as fundamental assumptions may not be satisfied. This will become clearer in the following chapters.

Table for converting ρ into r $\left(r = 2 \sin \frac{\pi}{6} \rho\right)$*.

ρ	r	ρ	r	ρ	r	ρ	r
·05	052	30	313	55	568	80	813
10	105	·35	364	60	·618	85	·861
15	157	40	416	·65	·668	·90	908
20	209	45	467	·70	717	·95	954
·25	261	50	518	·75	·765	1 00	1 000

* Quoted from K Pearson, F R S., *Drapers' Company Research Memoirs*, Biometric Series, iv 1907, p 18

CHAPTER VI

THE MATHEMATICAL THEORY OF CORRELATION

Correlation coefficient r—Correlation ratio η—Probable errors—The normal correlation surface and its properties—Other methods of determining correlation—Fourfold table—Method of contingency—Two-row table—Short methods—The method of ranks—Spearman's foot-rule—Correlation of sums or differences—Reliability Coefficients.

IN the present chapter an attempt will be made to summarise briefly the principal methods in use for obtaining a measure of the correlation, or tendency towards concomitant variation, of two or more variates

Let the coordinates of the dots in the accompanying diagram—commonly known as a "scatter diagram"—represent the measures of two separate characteristics, e.g speed of adding figures (x) and accuracy of adding figures (y), in a number of individuals (N).

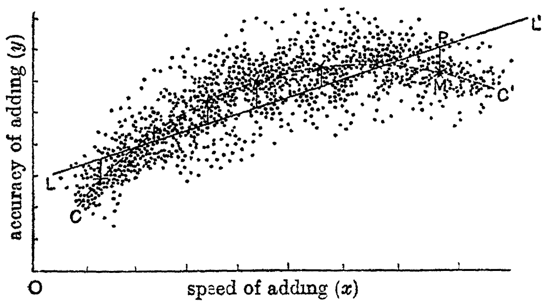


Fig 17

Let the crosses represent the mean values of y corresponding to values of x lying between the limits of pairs of successive units of measurement. Then the broken curve CC' passing through these crosses represents the most probable law of relationship between speed of adding and accuracy of adding, and is known as the *regression curve.* (In practice the crosses do not lie so accurately on the curve)

(1) CORRELATION COEFFICIENT (r)

What we chiefly want to know, however, even when the regression, as here, is not linear, is (1) whether large x is on the whole associated with large y, etc , (2) how to find roughly the mean y associated with given x. To do this, we find the "best fitting" straight line, LL', to the

swarm of dots in the figure, using, merely from motives of convenience, the Method of Least Squares*.

Let the equation of LL' be

$$Y = b_{21}X + c.$$

Then applying Least Squares will make $S(y - Y)^2$ a minimum, where y is the ordinate of any dot, and Y the ordinate of the line at the same abscissa, which is both X and x There will be as many equations

$$(y - Y) = y - (b_{21}x + c) = v$$

as there are dots, and the v's correspond to the "residuals" of the Method of Least Squares, though here they are real deviations and not errors The Normal Equations for b_{21} and c, formed according to the rule on p. 45, give at once

$$S(xy) - b_{21}S(x^2) - cS(x) = 0,$$
$$S(y) \;\; - b_{21}S(x) \;\; - cS(1) = 0.$$

If there are N points, $S(1) = N$, and if the x's and y's are measured from the mean of the whole, $S(x) = S(y) = 0$. The equations then become

$$S(xy) = b_{21}S(x^2),$$
$$Nc = 0.$$

Thus c is zero, that is the line LL' goes through the mean. And b_{21}, which is the tangent of the slope of LL', is

$$b_{21} = \frac{S(xy)}{S(x^2)} = \frac{S(xy)}{N\sigma_1^2}.$$

The line LL' is known as the regression line, and b_{21} as the coefficient of regression of y on x. If we define r as

$$r = \frac{S(xy)}{N\sigma_1\sigma_2},$$

then

$$b_{21} = r\frac{\sigma_2}{\sigma_1},$$

and the equation to LL' is

$$Y = r\frac{\sigma_2}{\sigma_1}X,$$

x and y being measured from their mean values.

An analogous equation

$$X = r\frac{\sigma_1}{\sigma_2}Y$$

gives the regression of x on y. There are thus *two* regression lines. If

* G Udny Yule, "On the Significance of Bravais' Formulae for Skew Correlation," *Proc. Roy Soc.* 1896, LX pp 477—489

x and y, in addition to being measured from their means, are also measured in terms of their standard deviations as unity, the regression equations become

$$X = rY \quad \text{and} \quad Y = rX,$$

and r is then itself the coefficient of regression of y on x and of x on y, the two regressions being equal.

Since $y - Y$ measures the distance in the y direction of any point from the regression line, the quantity

$$\frac{S\,(y - Y)^2}{N}$$

gives the mean square deviation, in the y direction, of all the points from the regression line,

$$= \frac{S\,(y - b_{21}x)^2}{N} = \frac{S\,(y^2)}{N} - 2b_{21}\frac{S\,(xy)}{N} + b_{21}{}^2\frac{S\,(x^2)}{N}$$

$$= \sigma_2{}^2 - 2\,r\frac{\sigma_2}{\sigma_1}.\sigma_1\sigma_2 r + r^2\frac{\sigma_2{}^2}{\sigma_1{}^2}\,\sigma_1{}^2$$

$$= \sigma_2{}^2\,(1 - r^2).$$

Hence the standard error or deviation made in estimating, by means of the regression equation, the value of y most probably associated with any particular x, is, on the average, given by

$$\sigma_2 \times \sqrt{(1 - r^2)},$$

and if the distribution is Normal it has this value not only on the average but for each array

r is known as the coefficient of correlation, and evidently must lie between the values $+ 1$ and $- 1$. If the regression line coincides with the regression curve, within the limits of errors of random sampling,—in other words, if the regression is linear—r is a measure of the degree of dependence between x and y. When $r = \pm 1$, the points close up upon the line and the "scatter diagram" contracts to become the line itself.

The formula
$$r = \frac{S\,(xy)}{N\sigma_1\sigma_2}$$

is implied in Bravais' work of 1846, and was shown by Professor Karl Pearson in 1896 to be the best measure of r. Hence it is known as the Bravais-Pearson Product-Moment Formula. It may be written

$$r = \frac{S\,(xy)}{\sqrt{S\,(x^2)}\,\sqrt{S\,(y^2)}},$$

the denominator being the geometrical mean of the two second moments, and the numerator the *product*-moment, of x and y.

If x and y are not measured from their means, but from some convenient point distant d_1 from the x mean and d_2 from the y mean, the arithmetic is very considerably lightened, and the formula becomes, as may be tested by simple algebra,

$$r = \frac{S(xy) - Nd_1d_2}{\sqrt{\{S(x^2) - Nd_1^2\}}\sqrt{\{S(y^2) - Nd_2^2\}}}.$$

The following example is intended to show one form of calculation based on this formula Being only a model for this purpose, it is kept short so that the arithmetic can be easily followed *But it must be made quite clear that really to calculate correlations with only ten cases is absurd,* for the probable errors are enormous and moreover are unknown (see p. 114). A calculation with larger numbers, and made by a slightly different process, is given on p 115 *et seq.*

In the following table A and B are the percentage errors made by certain cadets in a test in judging distance, in the years 1915 and 1916 respectively

Name	A	B	$x=(A-16)$	$y=(B-13)$	x^2	y^2	xy
Cadet A	15	15	-1	2	1	4	-2
„ B	19	10	3	-3	9	9	-9
„ C	11	14	-5	1	25	1	-5
„ D	17	16	1	3	1	9	3
„ E	8	18	-8	5	64	25	-40
„ F	14	12	-2	-1	4	1	2
„ G	24	14	8	1	64	1	8
„ H	28	11	12	-2	144	4	-24
„ K	7	8	-9	-5	81	25	45
„ L	14	14	-2	1	4	1	-2
Means	15 7	13 2		Sums	397	80	-24
Provisional	16	13		Subtract	0 9	0 4	$-$ 0 6 (i e. Ndd)
	$d_1 = -$ 3	$d_2 = 2$			396 1	79 6	-23 4

$$r = -\frac{23\ 4}{\sqrt{(396\ 1 \times 79\ 6)}} = -\cdot 13.$$

(2) CORRELATION RATIO (η)*

It is clear that if the regression is not linear r ceases to be a satisfactory measure of the relation between the two characters under consideration. In an extreme case, such as that shown in the accompanying diagram, r may be zero while there is yet a very close relation between the two characters.

Clearly, if the individual observations, i.e. the dots in the figure, are

* See *Drapers' Company Research Memoirs,* Biometric Series, II p. 9 *et seq* Karl Pearson, "On the Theory of Skew-Correlation and Non-Linear Regression"

all exactly situated on the regression curve, the quantity y is an exact mathematical function of x, and correlation is perfect, or $\eta = 1$, where η is a new and as yet undefined measure of correlation, called the correlation ratio, while if the individual observations are much scattered right and left of anyone walking along the regression curve, the correlation is imperfect, and $\eta < 1$

If there is no scatter at all in any array, then the correlation is perfect, and the greater the scatter the less the correlation, i e the less certain is any prediction of y from x.

Professor Pearson therefore makes the correlation ratio η depend on the amount of scatter in the arrays Exactly, he makes it depend on the mean of the weighted squares of the standard deviations of the arrays, i.e. upon

$$\sigma_{a_y}^2 = S\,(n_x \sigma_{n_x}^2)/N.$$

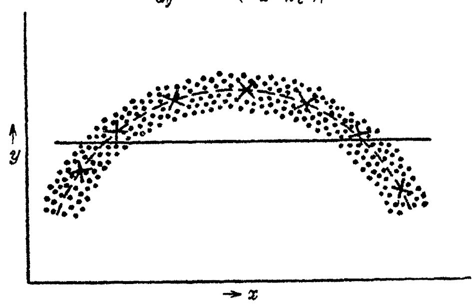


Fig. 18

The correlation ratio rises and falls as this quantity falls and rises n_x is the number of cases in the array of which σ_{n_x} is the standard deviation

Clearly the unit in which $\sigma_{a_y}^2$ is measured for this purpose must depend on the standard deviation of the whole of the y's. If the arrays are less scattered than the whole, there is correlation If there is no correlation, any array will have just as much scatter as the whole has Pearson writes

$$\frac{\sigma_{a_y}^2}{\sigma_2^2} = 1 - \eta^2.$$

Compare this equation with that found above for the case of linear regression, p. 109, namely

"average" mean square deviation of an array $= \sigma_2^2\,(1 - r^2)$,

which in our present notation is

$$\frac{\sigma_{a_y}^2}{\sigma_2^2} = 1 - r^2,$$

and from the comparison we see that Pearson has chosen η^2 so that it becomes r^2 for linear regression. The correlation ratio η however never becomes negative but is in linear regression numerically equal to r.

The above formulae already allow of the calculation of η, but a simplification is possible. By its definition,

$$\sigma_{n_x}{}^2 = S' (y_x - \bar{y}_x)^2 / n_x,$$

S' being a summation *up and down an array*, so that

$$\sigma_{a_y}{}^2 = SS' (y_x - \bar{y}_x)^2 / N,$$

S being a summation at right angles to the former, i.e. a summation of the arrays That is to say, the mean of the weighted squares of the standard deviations of the arrays is the same thing as the mean square of the distances of the dots (measured in the y direction) from the regression curve. This simplifies to

$$\begin{aligned} N\sigma_{a_y}{}^2 &= SS' (y_x{}^2) - 2SS' (y_x \bar{y}_x) + SS' (\bar{y}_x{}^2) \\ &= N\sigma_2{}^2 - 2S \{\bar{y}_x S' (y_x)\} + S (n_x \bar{y}_x{}^2) \\ &= N\sigma_2{}^2 - 2S (\bar{y}_x n_x \bar{y}_x) + S (n_x \bar{y}_x{}^2) \\ &= N\sigma_2{}^2 - S (n_x \bar{y}_x{}^2) \\ &= N\sigma_2{}^2 - N\Sigma_2{}^2, \end{aligned}$$

where Σ_2 is the standard deviation of the means of the arrays, each array being weighted with the number of cases in it. Therefore

$$\sigma_{a_y}{}^2 = \sigma_2{}^2 - \Sigma_2{}^2,$$

$$1 - \eta^2 = \frac{\sigma_{a_y}{}^2}{\sigma_2{}^2} = \frac{(\sigma_2{}^2 - \Sigma_2{}^2)}{\sigma_2{}^2} = 1 - \frac{\Sigma_2{}^2}{\sigma_2{}^2},$$

$$\eta = \frac{\Sigma_2}{\sigma_2}.$$

The correlation ratio, therefore, is the ratio of two standard deviations, one of the means of the arrays (properly weighted), the other of the whole.

Starting from the other variate, we arrive in a similar way at a second value

$$\eta = \frac{\Sigma_1}{\sigma_1}.$$

Since η is the ratio of two standard deviations, it must always be positive.

Let Y be the ordinate of any point on the regression *line*, then the average of the sum of the weighted squares of the distances between the regression line and the regression curve

$$= \frac{S\{n_x (\bar{y}_x - Y)^2\}}{N}$$

which reduces to $\sigma_2{}^2 (\eta^2 - r^2)$.

Thus η must always be numerically greater than r, except in the case of linear regression, when it is numerically equal to r.

In examining the relationship between two measurable characters, η should be calculated as well as r, since it serves as a test of the linearity or non-linearity of the regression, and is also a better measure of causal relation than r.

A simple criterion for linearity which is very generally applicable is that

$$\frac{\sqrt{N}}{\cdot 67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2} < 2{\cdot}5^*.$$

For very exact work, more complicated formulae need to be employed.

The results obtained above are all *independent of the forms of distribution* of the variates.

(3) PROBABLE ERRORS

In determining means, standard deviations, and other frequency constants, the investigator is unable to work from the "total population" and must be content with the results obtained from "random samples" of greater or less size taken from this (in some cases, hypothetical) total population

When the number of cases (n) in the random sample is fairly large—so large that fractions containing certain higher powers of n in the denominator can be neglected—the probable errors are found to be as follows†:

$$\text{P.E of a mean} = \cdot 67449 \frac{\sigma}{\sqrt{n}},$$

$$\text{,, ,,} \quad \sigma = \cdot 67449 \frac{\sigma}{\sqrt{2n}},$$

$$\text{,, ,,} \quad r = \cdot 67449 \frac{1 - r^2}{\sqrt{n}}.$$

The second and third of these values are only correct when the frequency-distribution is normal or approximately normal In particular, for large values of r the true P.E. may be considerably different from that given by the above formula unless the distribution is normal.

P.E. of $\eta = \cdot 67449 \dfrac{1 - \eta^2}{\sqrt{n}}$, for linear regression, and also, as a rough

* J Blakeman, *Biometrika*, IV. pp 349, 350

† See W Gibson, "Tables for Facilitating the Computation of Probable Errors," *Biometrika*, IV. p. 385 *et seq.* and Pearson's *Tables*

measure, for cases of skew correlation If greater exactitude is needed in the latter cases, more complicated formulae have to be employed*.

Another frequency constant in common use is the *coefficient of variation* V, which $= \dfrac{100\sigma}{\text{mean}}$.

Its P.E. $= \cdot 67449 V \left\{1 + 2\left(\dfrac{V}{100}\right)^2\right\}^{\frac{1}{2}} \Big/ \sqrt{2n}$†.

As stated above, the values just given for the probable errors only apply in cases where n is fairly large In cases where n is so small that certain higher powers of its reciprocal cannot be neglected in comparison with the rest of the expressions involving them, the values cannot be used. For such cases no theoretical formulae have hitherto been devised.

An empirical investigation has however been made‡ on samples of 4, 8, and 30 cases, taken from a "total population" of 3000 pairs of measurements (height and left middle finger measurements of 3000 criminals, "real" correlation, ·66) From the results obtained it may be concluded that, although in the case of such small samples as 4 or 8 the ordinary formula for the probable error of r gives much too low a value, yet in the case of as many as 30, the formula applies with tolerable accuracy. We must, however, bear in mind that this result has only been proved (empirically) to hold in the single case when the actual correlation was ·66.

The calculation of the probable errors of means, standard deviations, coefficients of variation, and coefficients of correlation is very much facilitated by the use of Pearson's *Tables for Statisticians*, especially Tables V, VI, VII and VIII, calculated by members of the staff§ of the Biometric Laboratory, University College, London.

We give next an example of the evaluation of r and of η between speed of adding single digits and accuracy in doing so, the individuals measured being 86 boys between the ages of 11 and 12 years from two L.C C elementary schools.* The two groups could be thrown together for this purpose, since the means and standard deviations calculated from them separately were in very close agreement—well within the limits of the probable errors.

* Karl Pearson, *op. cit* (Biometric Series, II) p 19

† Calculated values for different values of n given in Gibson's Tables, see pp xxi and 18 of Pearson's *Tables*

‡ "Student": "The Probable Error of a Coefficient of Correlation," *Biometrika*, 1908—9, VI. p. 302.

§ Miss Winifred Gibson, Dr Raymond Pearl, T. Blakeman, Dr David Heron, Miss H Gertrude Jones, H. E. Soper.

86 boys aged 11—12 years.

Correlation between speed and accuracy in the addition of groups of 10 single digits. Two tests, of 5 minutes' duration each.

Correlation Table.

→ Speed of Addition (x)

↑ Accuracy of Addition (y)	*100—140*	*140—180*	*180—220*	*220—260*	*260—300*	*300—340*	*340—380*	*380—420*	Totals (n_y)	Means ($\bar{x}_y$)
50—110	—	3 *11*	0 5 *5 5*	0 5 *0*	1 *5 5*	—	—	1 *22*	6	- 250
110—125	1 *9*	1 *6*	1 *3*	1 *0*	—	—	—	—	4	-1 500
125—140	—	2 *4*	0 5 *2*	0 5 *0*	—	—	—	—	3	-1 500
140—155	0 5 *3*	2 5 *2*	1 *1*	1 5 *0*	1 5 *1*	2 *2*	—	—	9	- 222
155—170	—	3 *0*	2 *0*	4 *0*	5 *0*	3 *0*	—	1 *0*	18	+ 389
170—185	1 *3*	4 5 *2*	4 5 *1*	5 5 *0*	5 5 *1*	3 *2*	0 5 *3*	0 5 *4*	25	- 060
185—200	1 *6*	2 5 *4*	5 *2*	5 *0*	2 5 *2*	2 5 *4*	2 *6*	0 5 *8*	21	+ 166
Totals (n_x)	3 5	18 5	14 5	18	15 5	10 5	2 5	3	N=86	$\bar{x}$= - 0698 σ_x^2= 2 796
Means ($\bar{y}_x$)	- 143	-1 000	+ 569	+ 403	+ 226	+ 571	+1 800	-1 333	$\bar{y}$= 081 σ_y^2=4 034	

Note. The figures in italics immediately beneath the frequency values within the correlation table are for the calculation of $S(xy)$. The row and column with zeros correspond to the arbitrary means from which the true means, S D 's and $S(xy)$ are calculated

Frequency	x'	Frequency $\times x'$	Frequency $\times x'^2$	Frequency	y'	Frequency $\times y'$	Frequency $\times y'^2$
3·5	−3	10·5	31·5	6	−5·5	33	181·5
18·5	−2	37	74	4	−3	12	36
14·5	−1	14·5	14·5	3	−2	6	12
18	0	−62	—	9	−1	9	9
15·5	1	15·5	15·5	18	0	−60	—
10·5	2	21	42	25	1	25	25
2·5	3	7·5	22·5	21	2	42	84
3	4	12	48	$86=N$		+67	347·5
$86=N$		+56	248			+7	
		−6					

$$d_1 = -\frac{6}{86} = -0{\cdot}06977, \qquad d_2 = \frac{7}{86} = 0{\cdot}0814,$$

$$\sigma_1{}^2 = \frac{248}{86} - (d_1)^2 - \frac{1}{12}*, \qquad \sigma_2{}^2 = \frac{347{\cdot}5}{86} - d_2{}^2\dagger,$$

$$= 2{\cdot}7955, \qquad = 4{\cdot}0341,$$

$$\therefore\ \sigma_1 = 1{\cdot}67 \qquad \therefore\ \sigma_2 = 2{\cdot}013.$$

Frequencies	$x'y'$‡	Total frequency f	$f \times x'y'$	
1+5·5−4·5−1·5	1	0·5	+0·5	
2·5+0·5+2·5+3−4·5−5−2	2	−3		−6
1+0·5+0·5−1	3	1	3	
2+2·5+0·5−2·5	4	2·5	10	
0·5−1	5·5	−0·5		−2·75
1+2−1	6	2	12	
0·5	8	0·5	4	
1	9	1	9	
3	11	3	33	
−1	22	−1		−22
			71·5−30·75	
			$S(x'y')=40{\cdot}75$	

* Sheppard's correction Remember that $\sigma^2=\mu_2$, and $d=\nu_1'$; see above, p. 84.

† Sheppard's correction cannot be used here, since the units of the subgroups are not equal and there is not high contact at the ends of the frequencies

‡ The figures in italics in the correlation table.

$$S(xy) = S(x'y') - Nd_1d_2$$
$$= 40\ 75 + 86 \times \cdot 0057$$
$$= 41\cdot 24,$$
$$r = \frac{S(xy)}{N\sigma_1\sigma_2} = \frac{41\cdot 24}{86 \times 1\ 67 \times 2\ 013} = 0\cdot 143,$$
$$\text{P E} = \cdot 67449 \cdot \frac{1 - r^2}{\sqrt{N}} = 0\ 071.$$
$$\therefore\ r_{\substack{\text{speed of addition}\\ \text{acc of addition}}} = \mathbf{0\cdot 14 \pm 0\cdot 07}.$$

$\bar{y}_x - \bar{y}$	$(\bar{y}_x - \bar{y})^2 \times n_x$
−0 224	050176 × 3 5 = 175616
−1 081	1 168561 × 18 5 = 21 618379
0 488	·238144 × 14 5 = 3 453088
0 322	103684 × 18 = 1 866312
0 145	·021025 × 15 5 = 325888
0 490	·2401 × 10 5 = 2 52105
1 719	2 954961 × 2 5 = 7 387403
−1 414	1 999396 × 3 = 5 998188
	$S\{n_x(\bar{y}_x - \bar{y})\}^2$ = 43 345924

$$\eta^2 = \frac{S\{n_x(\bar{y}_x - \bar{y})^2\}}{N\sigma_y^2}$$
$$= \frac{43\cdot 345924}{86 \times 4\cdot 034} = 0\cdot 1249,$$
$$\therefore\ \eta = 0\ 353,$$
$$\text{P.E} = \cdot 67449\,(1 - \eta^2)/\sqrt{N}$$
$$= 0\cdot 064$$
$$\therefore\ \eta_{\substack{\text{speed of addition}\\ \text{acc of addition}}} = \mathbf{0\cdot 35 \pm 0\cdot 06}.$$

Calculating η from the means of the y-arrays, we have

$$\eta^2 = \frac{S\{n_y(\bar{x}_y - \bar{x})^2\}}{N\sigma_x^2}$$
$$= \frac{19\cdot 68101}{86 \times 2\cdot 796}$$
$$= \cdot 0819.$$
$$\therefore\ \eta = \mathbf{0\cdot 29 \pm 0\cdot 07}.$$

To test the value of η obtained from the means of the x-arrays, for linear regression

$$\frac{\sqrt{N}}{\cdot 67449} \cdot \tfrac{1}{2}\sqrt{\eta^2 - r^2}* = \frac{\cdot 323}{2 \times \cdot 07273}$$
$$= 2\cdot 22, \text{ i.e. } < 2\cdot 5.$$

Hence regression may be considered to be linear.

* See p. 113.

Regression coefficients:

$$b_{12} = r\frac{\sigma_1}{\sigma_2} = \cdot 118,$$

$$b_{21} = r\frac{\sigma_2}{\sigma_1} = \cdot 172.$$

Equations to regression lines are

$$x - \bar{x} = b_{12}(y - \bar{y}),$$

and $$y - \bar{y} = b_{21}(x - \bar{x}).$$

$\therefore$ equation to regression line AB is

$$y - 164{\cdot}2 = \cdot 172\,(x - 237\ 7),$$

i.e. $$y = \cdot 172x + 123\ 316 \qquad \ldots\ldots \text{(i)}.$$

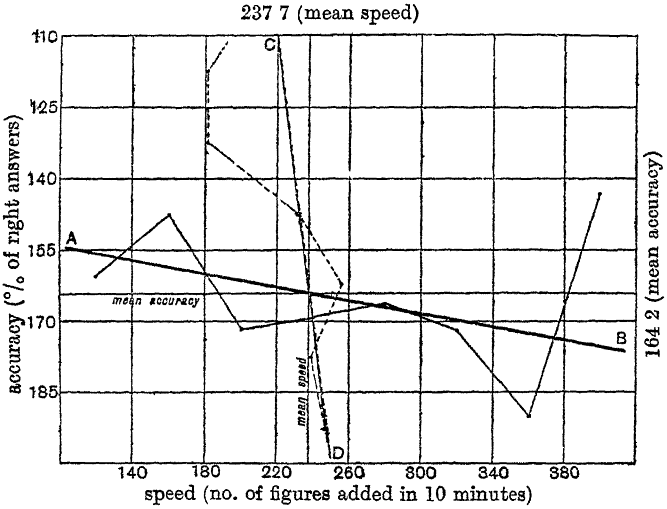


Fig 19

Similarly the equation to the line CD is

$$x - 237{\cdot}7 = \cdot 118\,(y - 164{\cdot}2),$$

i.e. $$x = \cdot 118y - 218{\cdot}324 \qquad \ldots\ldots \text{(ii)}.$$

Equation (i) gives the most probable value of y associated with a given value of x, with a standard error

$$\sigma_2\sqrt{(1 - r^2)}, \text{ i.e. } 1{\cdot}991.$$

Similarly, *mutatis mutandis*, with equation (ii).

As a model of the proper use of the correlation coefficient and ratio we may cite a very detailed investigation into the relationship of intelligence to the size and shape of the head, and to other physical and

mental characters, by Prof. Karl Pearson*, which appeared in *Biometrika*, v. 1906—1907. The subjects measured were 1000 Cambridge undergraduates and considerably more than 5000 school-children Special care was taken in drawing up a quantitative scale of intelligence, adjustments being made so that the results fitted a "normal" or Gaussian distribution. The correlations were worked out in several different ways, —correlation coefficient (r), correlation ratio (η), coefficient of mean square contingency†, and the method, first suggested in this paper, of the *analograph*. Pearson describes this last method as follows· "In the case of intelligence, I take a normal scale as my base line and plot up the *percentage* of the character for each grade of intelligence along the centroid vertical of the corresponding range, drawing a horizontal line to represent the mean percentage in the population at large We thus obtain a diagram, which I will venture to call an *analograph*

"If the percentage increases *or decreases* continually with intelligence (or with the base character, whatever it may be), I term the relationship *homoclinal*, if the percentage does not reach its maximum with the maximum or minimum of intelligence, I term the diagram *heteroclinal*."

(4) THE NORMAL CORRELATION SURFACE AND ITS PROPERTIES

We have so far considered the correlation coefficient r from the point of view from which it was approached by Galton, who measured it by the inclinations θ_1 and θ_2 of what we have called the regression lines, according to the formulae

$$r = \frac{\sigma_1}{\sigma_2} \tan \theta_1 = \frac{\sigma_2}{\sigma_1} \tan \theta_2 = \sqrt{(\tan \theta_1 \tan \theta_2)}.$$

We followed this up by an application of the Method of Least Squares, first made by Mr Udny Yule, obtaining the more advantageous formula

$$r = \frac{S\,(xy)}{N\sigma_1\sigma_2},$$

known as the Bravais-Pearson Product-Moment Formula. All we have so far done is independent of the form of distribution of the correlated variates.

From a historical point of view however this is not quite the way in which the subject developed. In 1846 A Bravais published, in Vol IX of the *Mémoires de l'Institut de France*, an article entitled "Analyse mathématique sur les probabilités des erreurs de situation d'un

* Karl Pearson: "On the Relationship of Intelligence to Size and Shape of Head, and to other Physical and Mental Characters," *Biometrika*, 1906—1907, v pp 105—146.

† See below

point" This article does not in any sense deal with the modern idea of correlation, though the word is casually used once on p 263. Its mathematics however is applicable to the correlation problem.

Galton's work was done in ignorance of this article, and of course was work applied directly to certain social and anthropological problems, whereas Bravais' is a piece of mathematics only. Also in ignorance of Bravais' work, Professor F Y. Edgeworth* took some important steps on the road later followed by Professor Karl Pearson, who connected Galton's work with Bravais', adopted the product-moment formula which was implicit in the latter's equations, and showed that it is the best formula for the purpose, i e. it has the least probable error†.

Mr Udny Yule still later employed the simple method of arriving at this formula which we have used, in an article the chief importance of which is that it points out the fact, then first adequately realised, that the product-moment formula has a definite significance even if the distribution of errors is not normal

In the present section we shall now consider more definitely *normal* correlation.

If the x variate is normally distributed according to the law

$$\frac{1}{\sqrt{(2\pi)}\,\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \quad \ldots\ldots (1),$$

and the y variate according to the law

$$\frac{1}{\sqrt{(2\pi)}\,\sigma_2} e^{-\frac{y^2}{2\sigma_2^2}} \quad \ldots\ldots (2),$$

then the probability of simultaneous occurrence of a value x and a value y is

$$P = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}\right)} dx\,dy \quad \ldots\ldots (3),$$

provided x and y are independent or uncorrelated. If however they are correlated, this probability is

$$P' = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-r^2)}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2}-\frac{2rxy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)} dx\,dy \quad \ldots\ldots (4),$$

which reduces to the former expression when r, the coefficient of correlation, becomes zero. The surface

$$z = P'/dx\,dy \quad \ldots\ldots (5)$$

* *Phil Mag* 1892 and 1893, several articles In a paragraph buried in one of these, indeed, Professor Edgeworth reached, but did not realise the importance of, the product-moment formula, which he there gives as the best formula (*Phil. Mag* July 1893, Series 5, XXXVI on p. 100)

† See *Biometrika*, 1920, XIII. p. 25, for Pearson's views of Bravais' work.

is the normal correlation surface, and is a hillock shaped like a bell with an oval mouth

When two variates are recorded on a grid-iron table like that used in the above example of Speed of Addition and Accuracy of Addition, the resulting table is called a "correlation table." As, owing to the experimental difficulties of the subject, psychological correlation tables are seldom smooth enough to illustrate vividly the points about to be mentioned, we quote here a correlation table from the study of heredity which has already been used elsewhere for a similar purpose*

Father's Stature

Son's Stature	Inches	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
	60					2	2	4										
	61					2				4								
	62		1	1		2	4	1	1	2	2							
	63	.	1	1	9	9	8	16	20	11	5		1	1				
	64	4	.	6	15	12	17	32	37	12	5	6	3	5				
	65	8	4	2	8	13	38	*54*	*43*	30	22	14	10					
	66	.	2	4	9	21	38	*40*	*67*	*70*	*64*	21	8	10	4			
	67		6	8	19	14	*55*	*79*	*106*	*103*	*78*	*50*	*55*	13	2	4		
	68			6	8	30	*40*	*41*	*97*	*126*	*94*	*118*	*53*	34	38	9		
	69			4	.	21	20	*51*	*73*	*64*	*96*	*116*	*86*	*40*	14	9		4
	70					4	10	23	*75*	*47*	*78*	*90*	*78*	*58*	25	14	6	4
	71						13	20	35	*43*	*76*	*59*	*83*	*43*	32	20	4	4
	72						1	12	5	28	31	*43*	*45*	*40*	34	11	2	
	73							3	3	10	30	26	24	30	25	13	2	2
	74					4		6	6		21	9	10	26	13	13		8
	75										4	8		10	3	7	2	
	76										5	1		2	4	4		
	77										5	1	4			6		
	78											4	4		1	3		
	79														1	1		

For simplicity in printing, the numbers in the original table in *Biometrika*, 1902—3, II p. 415, have been multiplied by four this eliminates quarters and halves which occur through some heights being half-way between whole inches.

* E g. Mr Udny Yule's text-book on the Theory of Statistics, and, independently, by G. H. Thomson, "Mathematics and the Inductive Methods of Logic," *Proc Univ Durham Phil Soc.* 1912—13, v pp 76—99

It is interesting to look at a correlation table in greater detail If we think of it as a plane horizontal surface, and erect over the centre of each compartment a vertical line proportional to the number written in that compartment, then the tops of these lines touch the correlation surface.

It is clear from the equation to the surface that the contours, or lines of equal z, are ellipses, given by the equation

$$\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} = \text{constant} \qquad .(6).$$

In the figure the numbers 40 and over have been printed in italics so that this contour line can be approximately followed. It is seen to be roughly elliptical, the major axis of the ellipse lying obliquely. The major axes of all the contour ellipses of a surface showing correlation are inclined to the axes of coordinates, and if, as we always can, we choose the linear units of x and y on the diagram in such a ratio that

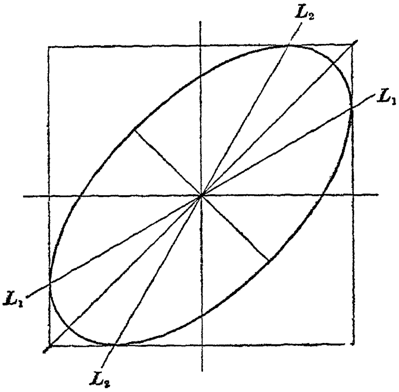


Fig. 20

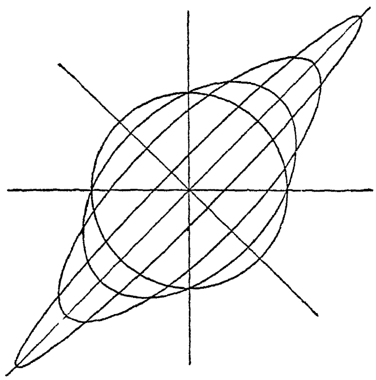


Fig. 21

$\sigma_1 = \sigma_2$, then this inclination will be at 45°, as shown in the accompanying figure (Fig. 20).

In the case of no correlation, $r = 0$, and if $\sigma_1 = \sigma_2$ the equation (6) given above for the contour lines represents a circle In this case, therefore, with suitable units of x and y, the contours are circles, and the correlation surface is a perfectly symmetrical hillock. As correlation increases, the contour lines become more and more drawn out along the 45° line (or the 135° line for negative correlation), as suggested in the figure (Fig. 21).

In a contoured plan of a correlation surface, those lines through the origin are important which cross all contour lines at points where the

tangents to the contour lines are parallel to the axes of coordinates Such are L_1L_1 and L_2L_2 in Fig. 20 If we differentiate equation (6) with regard to x and equate to zero we shall obtain L_2L_2. We find

$$\frac{2x}{\sigma_1^2} - \frac{2ry}{\sigma_1\sigma_2} = 0,$$

or simplifying

$$\frac{x}{\sigma_1} = r\frac{y}{\sigma_2}.$$

Similarly the equation of L_1L_1 is

$$\frac{y}{\sigma_2} = r\frac{x}{\sigma_1}.$$

That is, these are the regression lines

We have said that any horizontal section of a normal correlation surface is an ellipse Any vertical section, on the other hand, is a normal probability curve. Consider first a vertical section parallel to the x-axis. Write $y = c$ in equation (4)

After a little simple algebraical arrangement this then reduces to the form

$$z = \frac{e^{-\frac{c^2}{2\sigma_2^2}} \Big/ (\sigma_2\sqrt{2}\sqrt{\pi})}{\sqrt{(2\pi)}\,\sigma_1\sqrt{(1-r^2)}}\, e^{-\left(x - rc\frac{\sigma_1}{\sigma_2}\right)^2 \Big/ \{2\sigma_1^2(1-r^2)\}}$$

This is a normal curve, with area

$$e^{-\frac{c^2}{2\sigma_2^2}} \Big/ (\sigma_2\sqrt{2}\sqrt{\pi}).$$

Its centre is at $x = rc\sigma_1/\sigma_2$, $y = c$, i e. on the regression line. Its standard deviation is $\sigma_1\sqrt{(1-r^2)}$ and is independent of c

Similar statements hold for a vertical section along a line $x = c'$, the constant standard deviation being $\sigma_2\sqrt{(1-r^2)}$

A similar procedure will show, with rather more cumbrous algebra, that *any* vertical section is a probability curve

(5) OTHER METHODS OF DETERMINING CORRELATION

1. Fourfold Table.

	x_1	x_2	
y_1	a	b	$a+b$
y_2	c	d	$c+d$
	$a+c$	$b+d$	N

(*a*) When the divisions pass through the means of both characters,

$$r = \sin\frac{\pi}{2}\frac{(a-b)}{(a+b)}.$$

This formula (Sheppard's) is of little practical use, since the mean values, in cases where the fourfold table is the only method which can be used, are generally unknown.

(*b*) Yule's coefficient of colligation*

$$\omega = \frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}.$$

This however is not equivalent to the correlation coefficient But if the divisions are not too far from the medians, $\frac{\sin \pi\omega}{2}$ is an approximation to r Burt has used it for Binet test elements where nothing but a twofold division is provided for†.

Taking the correlation table on p 121, for which product-moment $r = 0{\cdot}51$, various divisions may be tried. Taking $67\frac{1}{2}$ inches for fathers and $68\frac{1}{2}$ inches for sons (which approximate to the medians) we get

1425	729
581	1577

giving $\omega = 0{\cdot}394, \quad \sin\frac{\pi\omega}{2} = 0{\cdot}58.$

The probable error of ω is

$$0\ 67449\,\frac{1-\omega^2}{4}\sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}.$$

(*c*) Tetrachoric r. The correct value in such cases, provided a normal surface is assumed to underlie the fourfold table, is given with greatest probability by tetrachoric r. Let a be the quadrant in which the means lie, and write $N\tau_0 = b + d$, $N\tau_0' = c + d$, then r is obtained from

$$d/N = \tau_0\tau_0' + \tau_1\tau_1'r + \tau_2\tau_2'r^2 + \ldots + \tau_n\tau_n'r^n.$$

The values of $\tau_1, \tau_2 \ldots \tau_6$ corresponding to various values of τ_0 are given in Everitt's Tables of Tetrachoric Functions, *Biometrika*, VII p. 437, or Table XXIX in Pearson's *Tables*, where examples are given It is only occasionally needful to go beyond τ_6.

* Yule, *Journ. Roy Stat. Soc* 1912, LXXV p 592

† *Mental and Scholastic Tests*, London, 1921, p 217

Figure 22 omitted in this edition, cf. p. 92.

The probable error of r obtained by the fourfold table method is much larger than that given by the formula of p 113. The correct formula is too complicated to insert here*.

2 METHOD OF CONTINGENCY†

The following is an example of a contingency table‡

Fathers

		Merry	Melancholy	Alternating	Even	Totals
Sons	Merry	122	8	81	67	278
	Melancholy	10	2	7	10	29
	Alternating	70	9	101	68	248
	Even	58	6	66	45	175
	Totals ...	260	25	255	190	730

Arithmetically the method is as follows Divide the square of each of the above numbers by the product of the totals of its row and column. Thus for example

$$122^2 \div (278 \times 260) = \cdot 2056.$$

This gives the table·

2056	0092	0926	·0849
0131	0055	0066	·0179
0760	·0130	1614	·1010
0739	0082	0976	0609

The grand total of these fractions is 1·0274 and the coefficient of mean square contingency C_1 is

$$\sqrt{\frac{\cdot 0274}{1\cdot 0274}} = 0\ 16.$$

The method is employed when the grouping is merely by class and the different classes have no known relation to one another—in other words, when the grouping is merely qualitative. The order of the different qualities can be changed without making any difference to this method

* Unless the dichotomies are extreme, good values are obtained from the use of Tables XXIII and XXIV, Pearson's *Tables*

† Karl Pearson, "On the Theory of Contingency and its Relation to Association and Normal Correlation," *Drapers' Company Research Memoirs*, Biometric Series, I. 1904, Dulau and Co , London.

‡ Taken from a paper by E. Schuster and E M Elderton on "The Inheritance of Psychical Characters (being a further Statistical Treatment of Material Collected and Analysed by Messrs G. Heymans and E. Wiersma)," *Biometrika*, 1906—1907, V. pp. 460—469.

The relationship between the two variables is measured by the differences between the numbers actually found in the various compartments of the table, and the numbers that might be expected there by pure chance

To state the rule:

The total mean square contingency, ϕ^2, of the table is given by

$$\phi^2 = \frac{1}{N} S_{pq} \left\{ \frac{\left(n_{pq} - \dfrac{n_p n_q}{N}\right)^2}{\dfrac{n_p n_q}{N}} \right\}$$

where n_p = total frequency in pth row,
n_q = total frequency in qth column,
n_{pq} = frequency of constituent common to pth row and qth column,
N = total number of cases in the table.

Then the coefficient of mean square contingency C_1 is:

$$C_1 = \sqrt{\frac{\phi^2}{1+\phi^2}}*.$$

If it is assumed that a normal distribution underlies the classification, and if the fineness of grouping is right, then the coefficient C_1 is numerically equal to the correlation coefficient r.

In the above case, $C_1 = 0{\cdot}16$.

The probable error of C_1 is very complicated†. It may be taken as approximately one and a third that of r.

Instead of calculating the mean *square* contingency, it is easier though not so accurate to calculate the mean contingency. Each quantity

$$n_{pq} - \frac{n_p n_q}{N}$$

* There are certain corrections to ϕ^2, not mentioned here, which often make considerable difference to the result See K Pearson, F R S, "On the Influence of Broad Categories on Correlation," *Biometrika*, 1913. IX. pp 116—139 The important point is to have fine enough grouping, but not so fine as to leave cells with very few or no cases in them If there are κ columns and λ rows the assumption of a rectangular distribution (an unlikely assumption) leads to the corrective factor

$$\sqrt{\frac{\kappa\lambda}{(\kappa-1)(\lambda-1)}}$$

by which C_1 has to be multiplied. This correction is undoubtedly too large, and empirically the fourth (instead of the second) square root in the above factor gives better results

† J. Blakeman and Karl Pearson, "On the Probable Error of Mean Square Contingency," *Biometrika*, 1906, v. pp 191—197 A W. Young and Karl Pearson, "On the Probable Error of a Coefficient of Contingency without Approximations," *Biometrika*, 1916, XI pp 215—230.

is called a subcontingency, and it will be observed that in the formula for ϕ^2 these were squared. Instead of this, let us, without squaring them, add the *positive* subcontingencies only (for of course the sum of the whole is zero), and write

$$N\psi = S_{pq}\left\{n_{pq} - \frac{n_p n_q}{N}\right\}.$$

From ψ, by using Table XXXIV of Pearson's *Tables*, a value of the second coefficient of contingency C_2 is read off, which, under conditions similar to those outlined above, also is equivalent to r.

The contingency method, it must again be emphasised, gives these two measures of the connection or association of the qualities considered even without any assumption that a continuous variation underlies the discrete classification. If however such is assumed, then the approach to equality of C_1 and C_2 will be a good measure of the normality of the distribution and the suitability as to smallness of our elements of grouping With very fine grouping we get into difficulties owing to having to record by units only 16 to 25 subgroups is a good range

It is interesting to find that areas of positive contingency are separated from areas of negative contingency on a normal surface by a hyperbola having a simple relationship with the contour ellipses.

3. Two-row Table* (Biserial r).

This method gives a unique value of r in the case of two variates one of which is both quantitative and continuous (e g intelligence), while the other, though quantitative, admits of only two subdivisions (e g. into good and bad visualisers), or, in more technical language, is "alternative."

x (intelligence)

y	Good visualisers									pN
	Bad visualisers									qN
										N

* Karl Pearson, F.R S , *Biometrika*, 1909, VII. p. 97.

The assumptions made are two in number:

(1) that the regression is *linear*.

(2) that the distribution of the alternative variate is approximately normal or Gaussian.

The regression line
$$\frac{x}{\sigma_1} = r\frac{y}{\sigma_2}$$
must go through the centroids of the good visualisers and the bad visualisers The abscissae x' and x'' of these centroids can be found from the data They are the mean intelligence of the good visualisers and the mean intelligence of the bad visualisers respectively, measured from the total mean. The ordinates y' and y'' of the two centroids can be found from page 38. If there are pN good visualisers and qN bad visualisers then by that page
$$pNy' = \sigma_2{}^2 Z,$$
where Z is the ordinate of a normal curve divided into the areas pN and qN by Z, and having sigma equal to σ_2. Or
$$py' = \sigma_2 z,$$
where z is the corresponding ordinate of a normal curve of *unit* area and *unit* sigma given in Sheppard's Table (Table II of Pearson's *Tables*, the z corresponding to $p = \frac{1}{2}(1 + \alpha)$ of that table).

Similarly
$$qy'' = \sigma_2 z.$$

We have therefore
$$\frac{x'}{\sigma_1} = r\frac{y'}{\sigma_2}, \quad \frac{x''}{\sigma_1} = r\frac{y''}{\sigma_2},$$
$$\therefore \frac{x' + x''}{\sigma_1} = r\frac{y' + y''}{\sigma_2} = \left(\frac{z}{p} + \frac{z}{q}\right) r = zr\frac{p+q}{pq} = \frac{zr}{pq}.$$

$x' + x''$ is the distance between the centroids (since they are on opposite sides of the general mean from which x is measured) and therefore can be replaced by $m' - m''$, where m' and m'' are the means (measured from the ordinary zero) of the intelligence of good and bad visualisers. Therefore
$$r = \frac{m' - m''}{\sigma_1}\frac{pq}{z}.$$

Herein m' is the mean of the one class of whom there are pN cases, m'' the mean of the other class, z is the ordinate in Sheppard's Tables corresponding to $p = \frac{1}{2}(1 + \alpha)$ of those tables, σ_1 is the sigma of the continuous variate. The probable error of biserial r is approximately
$$0{\cdot}6745\,\frac{\sqrt{pq} - zr^2}{z\sqrt{N}}.$$

Figure 23 omitted in this edition, cf. p. 92.

In cases where the x-variate (continuous and quantitative in our example above) can only be divided into classes, showing no definite order or quantitative relations to one another, the y-variate being again quantitative and assumed to follow a normal distribution, but alternative, a modification of the above method gives η.

4. SHORT METHODS*.

(i) It can easily be shown that

$$\sigma^2_{x-y} = \sigma_x^2 + \sigma_y^2 - 2r_{xy}\sigma_x\sigma_y,$$

and therefore that

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma^2_{x-y}}{2\sigma_x\sigma_y}.$$

This formula has been ingeniously utilised by Toops† and by Otis to minimise the arithmetical work of product-moment r In a correlation table like that on p 121, σ_x and σ_y are of course obtained from the marginal totals of the rows and columns respectively. In identical manner σ_{x-y} can be obtained from the *diagonal* sums from low-low to high-high Along any such diagonal the difference between father's and son's stature is constant These authors have published‡ blank forms which direct the calculation and reduce errors to a minimum

(ii) If the distributions are both normal, and if both variates have the same mean, and the same standard deviation σ, then

$$r = 1 - \frac{\pi\{S(x-y)\}^2}{N^2\sigma^2},$$

where S is the sum of the *positive* differences only This method might sometimes be conveniently used in determining the "individual" correlation between performances of the same individuals in the same mental tests on different occasions

5. THE METHOD OF RANKS.

Some years ago the valuable suggestion was made by Professor C Spearman§ that measurements of psychical performance may con-

* Karl Pearson, "On Further Methods of Determining Correlation," *Drapers' Company Research Memoirs*, Biometric Series, IV 1907

† *Journ Exp Psychol.* 1921, IV p 434

‡ Toops, Teachers' College, Columbia Univ., New York Otis, World Book Co, Yonkers, N Y.

§ C Spearman, "Measurement of Association between Two Things," *Am Journ Psychol.* 1904, XV, "'Foot-rule' for Measuring Correlation," *Brit Journ of Psychology*, II Pt. I, July 1906 Were it possible to keep R in its place merely as a "foot rule" whose "chief mission is to gain quickly an approximate valuation of r," it would not be harmful But the ease of its calculation leads to its use too frequently "not merely for assay purposes as originally contemplated, but even sometimes for research" (see C. Spearman, *Brit Journ. of Psychol* 1910, III p 286).

veniently—nay, preferably—be replaced by the numbers representing the rank or order of merit of the individuals in the group On this basis the ordinary product-moment formula for r, $\frac{S(xy)}{N\sigma_1\sigma_2}$, can be easily shown to reduce to the form

$$\rho = 1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2 - 1)} \qquad \ldots\ldots (\alpha),$$

where ν_1 and ν_2 are the ranks of an individual in the two series.

Professor Spearman also suggested a still simpler formula, which he calls a "foot-rule" formula. It is

$$R = 1 - \frac{S(g)}{\frac{1}{6}(N^2 - 1)} \qquad \ldots\ldots (\beta),$$

where $S(g)$ denotes the sum of the "gains" in rank (sum of positive differences) of the second series on the first, and then empirically, by noting the distribution of a large number of chance values of $S(g)$,

$$r = \sin\left(\frac{\pi}{2} R\right) \qquad \ldots (\gamma).$$

This method of using ranks, and the formulae suggested therefor, were vigorously criticised by Karl Pearson in the paper quoted on the preceding page*.

Some form of frequency distribution must be assumed, and the "foot-rule" method assumes that form to be a *rectangle* On the assumption of normal distribution, Professor Pearson shows that

$$r = 2 \sin\left(\frac{\pi}{6}\rho\right) \qquad \ldots\ldots (\delta),$$

where ρ has the value given above

In terms of the sum of *positive* differences of ranks ("gains" in rank) the formula is

$$r = 2 \cos 2\pi \left\{\frac{S(g)}{N^2 - 1}\right\} - 1$$

$$= 2 \cos \frac{\pi}{3}(1 - R) - 1 \qquad \ldots\ldots (\epsilon).$$

This agrees very closely with Spearman's formula (γ), and, having a definite theoretical basis, should now take its place.

(6) CORRELATION OF SUMS OR DIFFERENCES†

This article is concerned with the following problem. After calculating the correlations between several series of values, it frequently

* Professor Spearman endeavours to meet some of these criticisms in *Brit. Journ. of Psychol.* 1910, III. p 271 *sqq.*

† C. Spearman, *Brit. Journ of Psychol* 1913, V. p. 417.

happens that we want the correlations given by some of the series added together, and differences are not less important than sums. The correlation of the pool is *not* the mean of the correlations. The general problem is as follows

Let the two series of values be denoted by $a_1, a_2 \ldots a_p$, and $b_1, b_2 \ldots b_q$, each being measured from its own mean and consisting of N cases Let these variates be multiplied by constants or weights. Required the correlation between

$$A = n_1a_1 + n_2a_2 + \ldots + n_pa_p,$$

and
$$B = m_1b_1 + m_2b_2 + \ldots + m_qb_q.$$

Since all the a's and b's are measured from their means, it is clear that A and B are also so measured. The required correlation is therefore

$$r = \frac{S'(AB)}{\sqrt{S'(A^2)}\sqrt{S'(B^2)}},$$

where the symbol S' indicates summation from 1 to N We shall retain the symbol S, on the other hand, for summation from 1 to p or 1 to q.

Consider now the correlation of any particular a with any particular b It is given by

$$N\sigma_a\sigma_b r_{ab} = S'(ab).$$

Multiplying the a by its constant n, and the b by its constant m, only alters the standard deviations of these quantities, not their means, since they are already measured from means We have therefore

$$Nnm\sigma_a\sigma_b r_{ab} = S'(na.mb),$$

and summing this over the p and q measurements of a and b we get

$$NS(nm\sigma_a\sigma_b r_{ab}) = SS'(na.mb).$$

Now the right-hand side of this equation means the sum of all possible products of na and mb. But a little consideration will show that this is exactly what the numerator of r, viz. $S'(AB)$, means. Therefore

$$S'(AB) = NS(nm\sigma_a\sigma_b r_{ab}).$$

The two quantities in the denominator of r are found similarly, or by putting $A = B$ in the expression just arrived at, and we have finally

$$r = \frac{S(nm\sigma_a\sigma_b r_{ab})}{\sqrt{\{S(nn\sigma_a\sigma_a r_{aa})\,S(mm\sigma_b\sigma_b r_{bb})\}}},$$

wherein the summations in the denominator include the correlations of an a or of a b with itself, correlations which of course are unity and each

case like $a_s a_t$ has a twin case $a_t a_s$. If all the σ's are equalised, if all the n's and m's are unity, and there are p a's and q b's, this becomes*

$$r = \frac{S\,(r_{ab})}{\sqrt{\{p + 2S\,(r_{aa})\}}\sqrt{\{q + 2S\,(r_{bb})\}}}.$$

(7) RELIABILITY COEFFICIENTS.

Anyone who has carried out psychological experiments, or even an ordinary examination, on the same subjects with similar tests on two or more different occasions does not need to be reminded that the results will differ, sometimes very decidedly. Unless however the differences are only slight, it is clear that the test or examination is of no practical use. Its reliability can be conveniently measured by the correlation coefficient of the marks obtained on the two different occasions Such a correlation coefficient is called a reliability coefficient In practice tests which give reliability coefficients lower than 0 7 are almost useless, and the ideal would be a great deal higher than this

In the case where we have two forms of a test correlating with one another r_1 and we wish to know how well a similar test p times as long would correlate with one q times as long, then with the assumptions involved in making in the above formula all r's $= r_1$, all σ's $= \sigma_1$ and all n's $=$ all m's $=$ unity, we obtain

$$r_{pq} = \frac{pqr_1}{\sqrt{\{p + (p^2 - p)\,r_1\}}\sqrt{\{q + (q^2 - q)\,r_1\}}}.$$

For $p = q$ this gives the reliability to be expected if a test is made p times longer, r_1 being the present reliability, viz.

$$\frac{pr_1}{1 + (p - 1)\,r_1},$$

sometimes called William Brown's Formula†. If r_1 is the correlation of

* Several convenient forms of this are given by Wynn Jones, *Brit Journ. of Psychol.* 1924, xv p 20

† Proved independently by Spearman and by William Brown on pp 290 and 299 respectively of *Brit Journ of Psychol* 1910, III. Pt 3. William Brown's simple proof from first principles is as follows· "If x_1, x_2, x_1', x_2' be two pairs of results (x denoting deviation from the mean value), we may assume that

$$\sigma_{x_1} = \sigma_{x_2} = \sigma_{x_1'} = \sigma_{x_2'} = \sigma_x \text{ (say)},$$

and that
$$S\,(x_1x_1') = S\,(x_1x_2') = S\,(x_2x_1') = S\,(x_2x_2') = n\,\sigma_x^2 r_1.$$

Hence we get
$$r_2 = \frac{S\,(x_1 + x_2)\,(x_1' + x_2')}{n\,\sigma_{x_1+x_2}\,\sigma_{x_1'+x_2'}}$$
$$= \frac{4n\sigma_x^2 r_1}{n\,(2\sigma_x^2 + 2r_1\sigma_x^2)}$$
$$= \frac{2r_1}{1 + r_1} \text{ Q E D.}$$

the two halves of a split test, then for $p = 2$ the last formula gives the reliability of the complete test. The assumptions made do not always hold, as Holzinger* has shown empirically.

It is easily seen that the amalgamation of four tests gives a reliability coefficient $= \dfrac{4r_1}{1+3r_1}$, and, in general, for p tests we have

$$r_p = \frac{pr_1}{1+(p-1)r_1}.$$

This last formula furnishes a ready means of determining from the reliability coefficient of a single test, the number of applications of the test which would be necessary to give an amalgamated result of any desired degree of reliability "

* *Journ of Educ. Psychol* 1923, xiv p 302 See also Crum, *Amer Math Monthly*, 1923, xxx p. 296 and a reply by Kelley, *Journ. of Educ Psychol* 1924, xv p 193

CHAPTER VII

THE INFLUENCE OF SELECTION

Influence of mild selection on σ and r—Rigorous selection and partial correlation—Three correlated variables represented by dice throws—Multiple correlation—Spurious correlation—Variate difference correlation method.

(1) THE INFLUENCE OF MILD SELECTION

THE essential point about the whole theory of correlation is that it tells us how a group of individuals selected from the general population according to some characteristic (say as being within certain limits of height, or possessing some mental ability or manual dexterity in a high degree) will also differ from the general population in other characteristics

The ordinary correlation coefficient already tells us much in this respect For example, if the correlation between two abilities, say (1) the ability, whatever it may be, which is measured by Dr McDougall's Dotting Machine and (2) the ability to memorise Nonsense Syllables according to certain experimental regulations, be known to be ·4 for the whole population, this means that if a group be selected with "Dotting" ability equal to x (measured from the general mean in σ units) then this group will most probably have an average "Nonsense Syllable" ability equal to $\cdot 4x$ (measured in a similar way).

Clearly in practice we do not usually know the means and the correlation for the whole population but only for samples. We take large samples and endeavour to ensure that they are random and not selected samples from the population we wish to investigate.

The selection contemplated in the above example is very rigorous: all the individuals are presumed alike in regard to "Dotting" ability. In practice such a rigorous selection never takes place. The boys in a school form, for instance, are more alike in say ability in Latin than the general population, yet not absolutely alike. The "scatter" of this variate (Latin) has been reduced, yet not to zero.

Just as selecting a group of individuals for one variate will alter the average value of other variates, so it will alter the scatter of these other variates, and their intercorrelations. It is this phenomenon which

in an extreme form gives us what we already know as "partial correlation" (see p 105).

In fact, selecting a group of individuals within certain limits of a quality A implies an indirect and less rigorous but frequently very important selection of the other qualities B, C, . of these individuals and of their intercorrelations Consider the simplest case of three organs, A being directly, B and C only indirectly selected. Let subscripts 1, 2 and 3 refer to A, B and C respectively, and let the standard deviations and correlations in the general population be σ_1, σ_2, σ_3, r_{12}, r_{23} and r_{31} In the selected group σ_1 is reduced by the selection to s_1, and σ_2 and σ_3 are indirectly altered to Σ_2 and Σ_3, r_{12}, r_{23}, r_{31} to $\mathbf{r}_{12}$, $\mathbf{r}_{23}$, $\mathbf{r}_{31}$. Then the following formulae enable these quantities to be calculated*.

Write
$$s_1/\sigma_1 = \cos\chi_1$$
Then
$$\Sigma_2/\sigma_2 = \sin a_{12}, \text{ and } \Sigma_3/\sigma_3 = \sin a_{13},$$
where
$$\cos a_{12} = r_{12}\sin\chi_1, \text{ and } \cos a_{13} = r_{13}\sin\chi_1.$$
Further
$$\mathbf{r}_{12} = \cot\chi_1 \cot a_{12},$$
$$\mathbf{r}_{13} = \cot\chi_1 \cot a_{13},$$
$$\mathbf{r}_{23} = \frac{r_{23} - \cos a_{12}\cos a_{13}}{\sin a_{12}\sin a_{13}}.$$

For example let us suppose for the moment that the correlations between (1) Classics, (2) Drawing, and (3) English have, in the general population of English Fourth Form boys, the values found on pp 104 and 105, viz.

$$r_{12} = \cdot 42,$$
$$r_{23} = \cdot 21,$$
$$r_{31} = \cdot 78,$$

and let us further suppose that, say on a standardised percentage system of marking, the standard deviations of the marks in these subjects are

$$\sigma_1 = 16,$$
$$\sigma_2 = 13,$$
$$\sigma_3 = 14.$$

Now suppose a mild selection of Fourth Form boys to be made on the basis of their ability in Classics, and in the selected group let us suppose that the standard deviation in marks in Classics is reduced to

$$s_1 = 12.$$

* See Karl Pearson, F R S , "On the Influence of Natural Selection on the Variability and Correlation of Organs," *Phil. Trans.* 1902, cc. A, pp. 1—66, where an interpretation in terms of spherical trigonometry is given.

On substituting these values in the above formulae we obtain the following tables:

Standard Deviations

	Before selection	Mild selection in Classics	Rigorous selection in Classics*
Classics	16	12	0
Drawing	13	12 5	11 8
English	14	12	8 7

Correlations

	Before selection	Mild selection in Classics	Rigorous selection in Classics*
Classics and Drawing	·42	·33	—
Classics and English	·78	·68	—
Drawing and English	21	·08	−·21

This mild selection for Classics has therefore left the scatter of ability in Drawing almost untouched, but has made the group somewhat more homogeneous than it was in English The intercorrelations are all slightly reduced, that between Drawing and English being now almost nil

These formulae show the result on r_{12} and r_{13} of a change in homogeneity made by reducing σ_1. They also show the effect on r_{23} of reducing σ_1, where the quantity (1) does not enter directly into the correlation. If in this last case the quantities (2) and (3) are equally correlated with (1) then we can in the above formulae write $a_{12} = a_{13}$ and we find

$$\mathbf{r}_{23} = \frac{r_{23} - \cos^2 a}{\sin^2 a},$$

whence

$$\frac{1 - \mathbf{r}_{23}}{1 - r_{23}} = \frac{\sigma_2{}^2}{\Sigma_2{}^2} = \frac{\sigma_3{}^2}{\Sigma_3{}^2}.$$

In words, when there is a change in homogeneity, unity minus the correlation coefficient is inversely proportional to the variance (a term suggested by Student, in *Biometrika*, 1923, for the square of sigma) This equation has been independently reached by Otis† for the purpose of correcting correlation coefficients measured in a group with narrow range, and by Kelley‡ for the special case of reliability coefficients It is only strictly applicable if the narrower range of the group has been produced by selection in a trait (1) which is (at least approximately) equally correlated with the two quantities (2) and (3) whose correlation is to be corrected: not if (2) or (3) have been directly selected for.

* See later.

† *Journ Educ Psychol* May 1922, p 293

‡ *Journ. Educ Research,* 1921, III. p 377.

(2) RIGOROUS SELECTION AND PARTIAL CORRELATION

If we suppose the selection in Classics to be absolutely rigorous, so that the resulting group is absolutely homogeneous in ability in that subject, then our formulae simplify considerably and we are left with

$$\begin{aligned} s_1 &= 0, \\ \therefore\ \chi_1 &= 90^\circ, \\ \sin\chi_1 &= 1, \\ \cos a_{12} &= r_{12}, \\ \cos a_{13} &= r_{13}, \\ \therefore\ \Sigma_2 &= \sigma_2\sqrt{(1-r_{12}^2)}, \\ \Sigma_3 &= \sigma_3\sqrt{(1-r_{13}^2)}, \end{aligned}$$

$\mathbf{r}_{12}$ is meaningless, the variate 1 being *fixed*, and similarly $\mathbf{r}_{13}$;

$$\mathbf{r}_{23} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{13}^2)}} = r_{23\cdot 1}.$$

This formula was reached by Pearson and Yule* before more general formulae were known, and is called the "partial" correlation of variates 2 and 3 for a constant value of 1, and is written $r_{23\cdot 1}$. Mr Yule obtained its value by applying to three variables the methods we have already, following him, employed on p. 108 for two, using the Method of Least Squares.

The p e. of a partial correlation coefficient is similar in form to that of a total correlation except that for the number of cases n we write $n - s$, where s is the number of variates eliminated†, i e.

$$0{\cdot}67449\,\frac{1-r^2}{\sqrt{(n-s)}}.$$

For the first partial correlation coefficient here discussed, $s = 1$ and

$$\text{P E of } r_{23\cdot 1} = 0{\cdot}67449\,\frac{1-r^2_{23\cdot 1}}{\sqrt{(n-1)}}.$$

Tables for calculating partial correlation have been given by Kelley‡, and graphic methods by Kelley§ and by E. R Wood‖

If we apply "partial" formulae to our three variables (1) Classics,

* K Pearson, *Proc Roy Soc* 1895, LVIII p. 241 (partial regression coefficients), and G. Udny Yule, "On the Significance of Bravais' Formulae for Skew Correlation," *ibid* 1896, LX pp 477—489

† R. A Fisher, *Metron*, 1924, III p. 329, and see Yule, *Proc Roy. Soc.* 1907, LXXIX. p. 182

‡ *Bulletin of the Univ of Texas*, May 10, 1916, No 27.

§ *Statistical Method*, New York, 1923, p 291.

‖ State Normal School, Emporia, Kansas

(2) Drawing, and (3) English, we shall get the standard deviations and correlations for a rigorous selection in Classics, given in the third column of the above tables. We see that the group is still heterogeneous in Drawing, but a good deal more homogeneous in English. The correlation between Drawing and English has now actually been reversed. Needless to say, the actual numbers in this example are not to be taken as giving the facts, being only used for the sake of illustrating the method

It is important to realise, and becomes clear from the above considerations, that *all correlations are partial correlations*, inasmuch as there is always a selection of the group we are working with, for age, or race, or social standing, or what not. Indeed even the whole living population is only a group, surviving from "what might have been," by natural selection. This wide point of view will save us from many of the errors into which we are apt to fall in handling correlation coefficients.

It is particularly tempting to draw what are usually fallacious conclusions from the comparison of "entire" and "partial" correlations as to the underlying factors at work causing the correlations For instance, in the present case one might be tempted to conclude that the original positive correlation between English and Drawing was entirely due to factors which these share with Classics, that is, to a general factor, and that any direct connection of these two subjects is of an "interference" nature. But such conclusions as to the underlying mechanism have to be made, if at all, with great reserve, as will be seen from examining cases where we have independent and first-hand knowledge of the factors at work, as we have for example in dice throwing A correlation can be set up between two dice throws of m and n dice respectively by leaving some of the m dice lying to form part of the second throw.

(3) THREE CORRELATED VARIABLES REPRESENTED BY DICE THROWS*

Let n red dice, n blue, n yellow, and n white dice be thrown, and let the variable x be given by the combined red and white, y by the combined yellow and white, and z by the combined blue and white scores, as in Fig. 24.

That is to say, there is a general factor (the white dice), common

* This section is an extract from an article by Godfrey H Thomson, *Brit Journ. of Psychol* 1919, IX p. 323 *et seq* Dice throws were used to illustrate a simple case of correlation by Weldon in *Lectures on the Method of Science*, Clarendon Press, 1906, p. 100 Brown in the first edition (1911) of this book, p 79, proves Weldon's formula theoretically. Thomson's proof of the more general formula on p 141 hereof was an extension of Brown's.

to all three variables, which causes all the correlations between them These correlations are

$$r_{xy} = r_{yz} = r_{zx} = \tfrac{1}{2}.$$

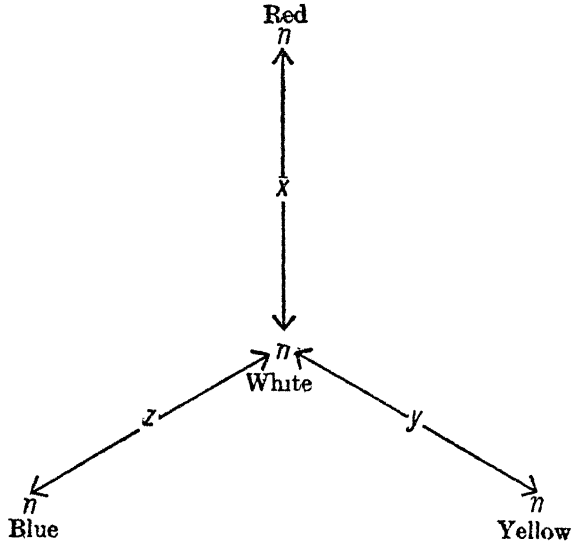


Fig. 24

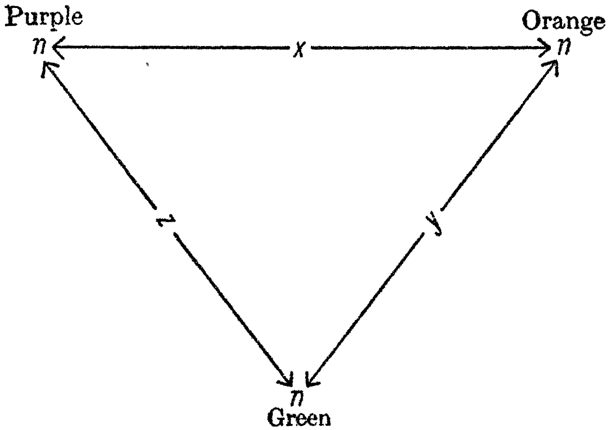


Fig. 25

The partial correlations are, by the well-known formula,

$$r_{xy\cdot z} = r_{yz\cdot x} = r_{zx\cdot y} = (\tfrac{1}{2} - \tfrac{1}{2}\cdot\tfrac{1}{2})/\sqrt{(1-(\tfrac{1}{2})^2)\,(1-(\tfrac{1}{2})^2)} = \tfrac{1}{3}.$$

It is not permissible, however, to reverse this statement, and to assume that in every case where $r_{xy} = r_{yz} = r_{zx} = \frac{1}{2}$ the correlations are formed

solely by the action of a general factor. In fact, identically the same values can be produced without any general factor at all Let n purple, n green, and n orange coloured dice be thrown, and let the variable x consist of the purple and orange, y of the orange and green, and z of the green and purple scores combined, as in Fig. 25.

Here there is no general factor whatever. The connection of x with y (through the orange dice) is entirely independent of the connection of x with z (through the purple dice).

Yet the correlations, both partial and entire, are exactly the same as in the first arrangement, viz.

$$r_{xy} = r_{yz} = r_{zx} = \tfrac{1}{2},$$
$$r_{xy.z} = r_{yz.x} = r_{zx.y} = \tfrac{1}{3}.$$

Clearly, therefore, if we only know of three variables x, y, and z, formed of dice throws, that their correlations are as above, we cannot say with certainty whether a general factor exists or not. Let us now consider a more general arrangement of dice, with numbers of different colours, viz. W white, R red, B blue, Y yellow, P purple, G green, and O orange, thus:

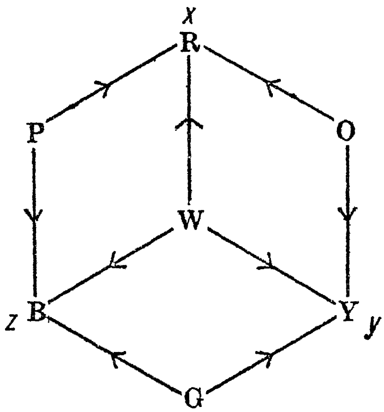


Fig. 26

x consisting of the scores of the W, R, P, O; y of the W, Y, O, G, and z of the W, B, G, P dice.

In this arrangement we shall call W a *general factor*, it being common to all three variables, R, Y, and B *specific factors*, they being unique to x, y and z respectively, and O, G and P *group factors*, since each runs through a group of (here two) variables.

The theoretical values of the correlation between any two of the variables, say x and y, can be found by means of the formula*

$$r_{xy} = \frac{\text{Number of dice common to } x \text{ and } y}{\text{Geometrical mean of total dice in } x \text{ and in } y}.$$

For example, if x is the score of $9n$ dice, and y the score of $4n$ dice, $2n$ being common, the correlation is

$$r = \frac{2n}{\sqrt{9n \times 4n}} = \frac{2}{6} = 0\ 33$$

The general arrangement of dice shown in the above figure includes the two special cases already considered, viz

$$(1)\ P = G = O = \text{zero},\quad R = B = Y = W = n,$$

and $$(2)\ P = G = O = n,\quad R = B = Y = W = \text{zero};$$

which both give $r_{xy} = r_{yz} = r_{zx} = \frac{1}{2}$†.

Of the infinite arrangements possible with this diagram, an infinite number (of a lower order) can in general be constructed to produce any given set of positive correlations between x, y and z‡. Moreover, all these possible ways of producing the required correlations are, in our ignorance, equally likely to have been those used by the person making the arrangement of dice, although they are, it is true, not equally probable as chance occurrences. From the correlations, therefore, we cannot in general deduce what proportion the white dice (i.e the particular colour representing the general factor) bears to the others, for this proportion can vary between wide limits, and give exactly the same correlations. The most we could conceivably do would be to give the

* This formula was proved by me in *Brit Journ of Psychol* 1916, VIII p. 275, in ignorance of any former statement of it. Professor Spearman showed (*ibid.* p 282) that it is deducible from a formula of his concerning correlation of sums or differences. I have since noticed that Professor Spearman gives the following clear expression of the formula (*Am. Journ. of Psychol* 1904, XV p 75): "The correlation is always the geometrical mean between the two shares." I do not however agree with the application he there proceeds to make The formula can also be directly deduced from Bravais, and in several other ways. It assumes that the elements all have the same standard deviation, as dice have. G H.T

† If he be so minded, the reader can make thousands of other dice patterns, all giving the correlations $r_{xy} = r_{yz} = r_{zx} = \frac{1}{2}$. One group is given by

$$R = B = Y = W = kn,\quad P = G = O = sn.$$

But such a high degree of symmetry is unnecessary. For example, another pattern giving the same correlations is

$$R = 91n,\ Y = 91n,\ B = 54n,\ P = 78n,\ G = 78n,\ O = 91n,\ W = 78n.$$

The number of white dice present ranges between the two extreme cases given in the text.

‡ A convenient method of finding such arrangements or patterns to give specified values of the correlation coefficients is explained by J. Ridley Thompson, *Brit. Journ. of Psychol.* 1919, x. p. 98.

"expectation" of the proportion of white dice. The meaning of this would be, that if a very large number of arrangements of dice were examined, each giving the required set of correlation coefficients, and *if we assume that these arrangements of dice are not formed on any plan,* beyond that they all agree in the correlations they produce, then the average proportion of white dice would be that named as the "expectation" thereof But if there is any reason to think that the large number of cases examined are all of much the same pattern—as there would be were they all natural phenomena of the same sort—then the "expectation" of the proportion of white dice becomes useless and meaningless. We cannot conclude anything which is of any definite value in constructing the pattern, except give limits within which it must lie.

This brings us to the problem —Are there any values of r_{xy}, r_{yz} and r_{zx} which make it certain (having regard to their probable errors) that at any rate *some* general factor, some number of white dice, exists? The answer is that this is so if the correlations are large enough. For example, if we take the special case of equality of the three coefficients,

$$r_{xy} = r_{yz} = r_{zx} = r,$$

then up to $r = \frac{1}{2}$ the correlations can be imitated by various numbers of red, blue, yellow, purple, green and orange dice without any white dice. But as soon as the common value r rises above $\frac{1}{2}$, some white dice are necessary. In this case therefore the proof of the existence of (at any rate) some amount of general factor reduces to the examination of the probable error of r, to see if r is indisputably greater than $\frac{1}{2}$.

In the more general case the matter is not so simple, the three values of r differing from one another. The more detailed examination of this case is reserved for treatment elsewhere. It will be found however that if the quantity

$$r_{xy}^2 + r_{yz}^2 + r_{zx}^2 + 2r_{xy}r_{yz}r_{zx}$$

is indisputably greater than unity, then some white dice, some general factor, may be postulated with certainty*. It may not be out of place to remind ourselves again that this, though true of the arrangements of dice we are considering, may not be true in the same sense of other phenomena, e.g. biological or mental phenomena.

The following two examples illustrate the above principles.

* A rough guide is the *average* value of the three r's: if this is greater than $\frac{1}{2}$ some general factor certainly exists. The exact condition however is that given in the text. It is due, in this form, to Mr J. R. Thompson, see *Brit. Journ of Psychol.* 1919, IX. p 335 Note that all this only applies to correlations produced by overlapping dice throws, or by some sufficiently similar mechanism.

Example A

Three variables, composed of overlapping dice throws, give correlations as follow·

$$r_{xy} = 0\ 32, \quad r_{yz} = 0{\cdot}33, \quad r_{zx} = 0\ 51.$$

Are any dice common to the three variables?

In this case we find

$$r_{xy}^2 + r_{yz}^2 + r_{zx}^2 + 2r_{xy}r_{yz}r_{zx} = 0\ 56, \quad \text{i e.} < 1$$

From this we conclude that these correlations can be imitated either with or without a general factor of white dice The following arrangements of dice do actually produce these correlations

Case (1)	$R = 19n$, $B = 17n$, $Y = 85n$,	(specific factors),
	$P = G = O =$ zero,	(no group factors),
	$W = 21n$,	(a general factor).
Case (2)	$R = 15n$, $B = 15n$, $Y = 16n$,	(specific factors),
	$P = 47n$, $G = 25n$, $O = 24n$,	(group factors),
	$W =$ zero,	(no general factor).

Example B

Three variables, composed of overlapping dice throws, give correlations as follow.

$$r_{xy} = 0{\cdot}72, \quad r_{yz} = 0{\cdot}77, \quad r_{zx} = 0\ 67.$$

Are any dice common to the three variables?

In this case we have

$$r_{xy}^2 + r_{yz}^2 + r_{zx}^2 + 2r_{xy}r_{yz}r_{zx} = 1{\cdot}93, \quad \text{i e} > 1.$$

We conclude therefore that some white dice common to the three variables are present, i.e that there is a general factor. The following arrangements of dice do actually produce these correlations, the general factor being a minimum in one and a maximum in the other

Case (1).	$R = 156n$, $B = 104n$, $Y = 52n$,	(specific factors),
	$P = G = O =$ zero,	(no group factors).
	$W = 260n$,	(general factor).
Case (2).	$R = B = Y =$ zero,	(no specific factors),
	$P = 90n$, $G = 161n$, $O = 123n$,	(group factors),
	$W = 198n$,	(general factor).

We see then that if

$$r_{xy}^2 + r_{yz}^2 + r_{zx}^2 + 2r_{xy}r_{yz}r_{zx} > 1,$$

the presence of some white dice is certain. If the above quantity, which

we shall call D, is equal to or less than unity, the presence of white dice is uncertain. Suppose we consider two cases in which

$$r_{xy} = 0{\cdot}8, \quad r_{yz} = 0{\cdot}4, \quad r_{zx} = 0{\cdot}1, \quad D = 0{\cdot}842,$$

and $$r_{xy} = 0{\cdot}2, \quad r_{yz} = 0{\cdot}2, \quad r_{zx} = 0{\cdot}1, \quad D = 0{\cdot}094, \text{ respectively.}$$

Can we in these two cases say anything as to the *probability* of the existence of some general factor?

The answer to this question is twofold. If we suppose that the person making the arrangements of dice has, among all the possible arrangements, chosen one by chance selection, then it is much more probable that a general factor exists in the first than in the second case This probability will in fact rise and fall with D though it is not *measured* by D But if the person making the arrangements of dice has any definite rules which he follows in making the patterns, then the above probability will have much less meaning

Before leaving for the present the subject of dice throws two points may be mentioned. (1) Negative correlations may be imitated by dice being added to one, but subtracted from the other, variable*. (2) There are many ways conceivable in which correlations can be produced other than by tangible common factors. Consider for example the positive correlation between the number of hearts in my hand and the number of spades in my partner's, at whist.

(4) MULTIPLE CORRELATION

For many purposes workers in experimental and educational psychology need not only the first partial correlation coefficient described on p 137, but also coefficients giving the correlation when more than one variable is kept constant. A very important case is when it is desired to weight a team of tests so as to correlate as highly as possible with some criterion, e g school success, or ability in a trade. This task can be carried out by finding all the intercorrelations of the tests with one another and with the criterion, and thence finding the regression equation in the manner explained below, with the criterion on the left and all the individual tests on the right The coefficients of that equation will indicate the way in which each individual test score should be weighted to give the highest correlation of the team with the criterion. Methods of shortening the arithmetic have been given by Toops† and by Chapman‡.

The classic memoir on the theory is that by Professor Karl Pearson

* See J. R Thompson, "The Rôle of Interference Factors in Producing Correlation," *Brit Journ of Psychol* 1919, x pp 81—100.

† *Journ. Educ Psychol* 1922, v p 68. ‡ *Ibid.* 1922, v. p. 263.

on "Regression, Panmixia and Heredity," in 1896*, which is however too advanced for quotation at any length here In 1907 Mr G Udny Yule introduced a new notation† and made various improvements, and his formulae will now be briefly summarised

If $x_1, x_2, \ldots x_n$ denote deviations from means, the equation expressing the regression of x_1 on $x_2 \ldots x_n$ is

$$x_1 = b_{12.34\ldots n}x_2 + b_{13.24\ldots n}x_3 + \ldots + b_{1n.23\ldots n-1}x_n,$$

where
$$b_{12.34\ldots n} = r_{12.34\ldots n}\frac{\sigma_{1.34\ldots n}}{\sigma_{2.34\ldots n}},$$

$$\sigma^2_{1.23\ldots n} = \sigma_1^2(1 - r_{12}^2)(1 - r^2_{13.2})(1 - r^2_{14.23}) \ldots (1 - r^2_{1n.23\ldots n-1}),$$

$$r_{12.34\ldots n} = \frac{r_{12.34\ldots n-1} - r_{1n.34\ldots n-1}r_{2n.34\ldots n-1}}{(1 - r^2_{1n.34\ldots n-1})^{\frac{1}{2}}(1 - r^2_{2n.34\ldots n-1})^{\frac{1}{2}}}.$$

$r_{12.34\ldots n}$ is known as a "partial" correlation coefficient, being the value of the correlation between 1 and 2 for constant values of $3, 4 \ldots n$ Similarly, $b_{12.34\ldots n}$ is a "partial" regression coefficient. Knowing the "total" correlations r_{12}, r_{13}, r_{23}, etc , we are enabled to obtain the various partial coefficients by successive substitutions. Thus, in the case of three variables, 1, 2, 3, we have

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}},$$

and two similar equations, expressing the value of the correlation between two of the variables for a constant value of the third‡. If a fourth variable be added, we have the further set of equations

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{(1 - r^2_{14.3})^{\frac{1}{2}}(1 - r^2_{24.3})^{\frac{1}{2}}}.$$

Perhaps a more convenient formula for obtaining the partial correlation of two variables for constant values of a third and fourth is

$$r_{12.34} = \frac{r_{12}(1 - r_{34}^2) - r_{13}(r_{23} - r_{24}r_{34}) - r_{14}(r_{24} - r_{23}r_{31})}{\sqrt{1 - r_{13}^2 - r_{14}^2 - r_{34}^2 + 2r_{13}r_{14}r_{34}}\sqrt{1 - r_{23}^2 - r_{34}^2 - r_{24}^2 + 2r_{23}r_{34}r_{24}}}.$$

When the partial correlation coefficients have been determined, the regressions can be found by substituting in the appropriate equations, and give at once the regression equations As explained before, a regression equation gives the most probable value of one variable for

* *Phil Trans* CXCVIII A, pp. 443—459

† *Proc Roy Soc* 1907, LXXIX A, pp. 182—193

‡ For an interesting representation of correlation between three variables by a model showing the distribution of points in space, see G Udny Yule, *An Introduction to the Theory of Statistics*, C. Griffin and Co., London, 1911, pp. 241—243.

given values of the remaining variables, the standard error in such a prediction being $\sigma_{1.23 \ldots n}$, etc The correlation between measured values of x_1 and the values of x_1 calculated, by means of the regression equation, from $x_2, x_3 \ldots x_n$ is $R_{1(23 \ldots n)}$, where*

$$1 - R^2_{1.(23 \ldots n)} = (1 - r_{12}^2)(1 - r^2_{13.2})(1 - r^2_{14.23}) \ldots (1 - r^2_{1n.23 \ldots n-1}).$$

Thus, for example, suppose that x_1 is proficiency in a trade after a year's experience, and x_2, x_3 and x_4 are three tests which were given before the year's work was entered upon Then in any future group tested, the best prediction as to trade ability a year hence will be obtained by combining the scores in x_2, x_3 and x_4 with weightings as in the regression equation

$$b_{12.34}x_2 + b_{13.24}x_3 + b_{14.23}x_4.$$

And this prediction may be expected to correlate with that future success to an extent $R_{1(234)}$ given by the equation just above†.

An example of the method of applying the above formulae is given in the next few paragraphs, where the partial correlations, regressions, etc in the case of four interrelated psychical capacities are worked out on lines identical with those illustrated by Mr Yule in his paper.

Example of Multiple Correlation‡

(*Boys, ages* 11—12, $n = 66$.)

1. Crossing through two letters (e and r).
2. Crossing through four letters (a, n, o, s).
3. Combination test
4. Mechanical memory test.

Formula:

$$r_{12.34 \ldots n} = \frac{r_{12.34 \ldots n-1} - r_{1n.34 \ldots n-1} r_{2n.34 \ldots n-1}}{(1 - r^2_{1n.34 \ldots n-1})^{\frac{1}{2}} (1 - r^2_{2n.34 \ldots n-1})^{\frac{1}{2}}}.$$

* It is of interest to note that the formula for R can be obtained from Spearman's formulae for the correlation of sums by making the latter a maximum (equating differentials to zero)

† See for example Toops, *Trade Tests in Education*, Teachers' College, Columbia University, 1921, and Clark L. Hull, *Journ Educ Psychol* 1923, xiv. p 396.

‡ W. Brown, *Brit. Journ. of Psychol.* 1910, iii p 317.

For four variables this becomes:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{(1 - r^2_{14.3})^{\frac{1}{2}}\,(1 - r^2_{24.3})^{\frac{1}{2}}},$$

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{(1 - r_{13}{}^2)^{\frac{1}{2}}\,(1 - r_{23}{}^2)^{\frac{1}{2}}}.$$

Table I

Correlation coefficient		$\log(1 - r^2)$
12	0·78	$\bar{1}$·59284
13	0·45	$\bar{1}$·90173
14	0·40	$\bar{1}$·92428
23	0·48	$\bar{1}$·88627
24	0·29	$\bar{1}$·96185
34	0·52	$\bar{1}$·86308

Table II

Correlation coefficient (zero order)		Product term of numerator	Numerator	Correlation coefficient (first order)		$\log(1 - r^2)$
12	0·78	0·2160	0·5640	12·3	0·7199	$\bar{1}$·68281
13	0·45	0·3744	0·0756	13·2	0·1377	$\bar{1}$·99169
23	0·48	0·3510	0·1290	23·1	0·2308	$\bar{1}$·97623
12	0·78	0·1160	0·6640	12·4	0·7570	$\bar{1}$·63038
14	0·40	0·2262	0·1738	14·2	0·2902	$\bar{1}$·96179
24	0·29	0·3120	−0·0220	24·1	−0·0386	$\bar{1}$·99348
13	0·45	0·2080	0·2420	13·4	0·3091	$\bar{1}$·95639
14	0·40	0·2340	0·1660	14·3	0·2176	$\bar{1}$·97893
34	0·52	0·1800	0·3400	34·1	0·4154	$\bar{1}$·91774
23	0·48	0·1508	0·3292	23·4	0·4027	$\bar{1}$·92316
24	0·29	0·2496	0·0404	24·3	0·0539	$\bar{1}$·99873
34	0·52	0·1392	0·3808	34·2	0·4536	$\bar{1}$·89996

Table III

Correlation coefficient (first order)		Product term of numerator	Numerator	Correlation coefficient (second order)		$\log(1 - r^2)$
12·4	0·7570	0·1245	0·6325	12·34	0·727	$\bar{1}$·67345
13·4	0·3091	0·3048	0·0043	13·24	0·007	$\bar{1}$·99998
23·4	0·4027	0·2340	0·1687	23·14	0·272	$\bar{1}$·96662
12·3	0·7199	0·0117	0·7082	12·34	0·727	—
14·3	0·2176	0·0388	0·1788	14·23	0·258	$\bar{1}$·97009
24·3	0·0539	0·1567	−0·1028	24·13	−0·152	$\bar{1}$·98985
13·2	0·1377	0·1316	0·0061	13·24	0·007	—
14·2	0·2902	0·0625	0·2277	14·23	0·258	—
34·2	0·4536	0·0400	0·4136	34·12	0·436	$\bar{1}$·90843
23·1	0·2308	−0·0160	0·2468	23·14	0·272	—
24·1	−0·0386	0·0959	−0·1345	24·13	−0·152	—
34·1	0·4154	−0·0089	0·4243	34·12	0·436	—

The regression equation between changes in intelligence (as measured by the combination test) and changes in the other three variables is

$$x_3 = b_{31.24}x_1 + b_{32.14}x_2 + b_{34.12}x_4.$$

Calculation of regression coefficients

$$\sigma_1 = 68{\cdot}19, \quad \sigma_2 = 58{\cdot}43, \quad \sigma_3 = 16{\cdot}34, \quad \sigma_4 = 9{\cdot}70,$$

$$\therefore\ \sigma_{3.24} = \sigma_3\,(1 - r_{32}{}^2)^{\frac{1}{2}}\,(1 - r^2{}_{34.2})^{\frac{1}{2}} = 12{\cdot}77.$$

Similarly,

$$\sigma_{1.24} = 40{\cdot}83,\ \sigma_{3.14} = 13{\cdot}27,\ \sigma_{2.14} = 36{\cdot}27,\ \sigma_{3.12} = 14{\cdot}23,\ \sigma_{4.12} = 8{\cdot}82,$$

whence $$b_{31.24} = r_{31.24}\frac{\sigma_{3.24}}{\sigma_{1.24}} = {\cdot}002.$$

Similarly, $\quad b_{32.14} = {\cdot}099, \quad b_{34.12} = {\cdot}703.$

Hence regression equation is

$$\mathbf{x_3 = {\cdot}002x_1 + {\cdot}099x_2 + {\cdot}703x_4.}$$

The standard error ($\sigma_{3.124}$) made in estimating x_3 from x_1, x_2 and x_4 by this equation

$$= \sigma_3\,(1 - r_{31}{}^2)^{\frac{1}{2}}\,(1 - r^2{}_{32.1})^{\frac{1}{2}}\,(1 - r^2{}_{34.12})^{\frac{1}{2}} = \mathbf{12{\cdot}78}$$

The multiple correlation $R_{3.(124)}$ of the measured x_3 with the x_3 estimated by this equation is given by

$$R^2{}_{3.(124)} = 1 - (1 - r^2{}_{31})\,(1 - r^2{}_{32.1})\,(1 - r^2{}_{34.12}),$$

$$\therefore\ R_{3.(124)} = \mathbf{0{\cdot}62.}$$

(5) SPURIOUS CORRELATION

Correlation is said to be *spurious* when it is due to extraneous conditions and does not arise directly out of the functions under consideration. The term is one of relative and not of absolute significance, but its appropriateness will become apparent after a consideration of the following two examples:

1. *Heterogeneity of material.*

Let us suppose that two distinct groups of children have been measured for two characters *A* and *B*, and that the mean abilities, in both *A* and *B*, are higher in the one group than in the other. Then, even if there is no correlation between the two characters, as estimated from each group separately, a positive correlation will be obtained from the two groups taken together. On the other hand, if the mean ability in *A* is higher, and the mean ability in *B* lower, in the one group than in the other, a negative correlation will be obtained by taking the two

groups together. The correlation in each case will be due simply to the *heterogeneity* of the material employed The difference in the mean values for the two groups must of course have some cause, such as a difference of nationality, sex, or even locality (within any one town or district) from which the children are drawn, or, again, such a cause as the difference of discipline to which the two groups have been subjected in the past These are extraneous conditions and, if measurable, can be allowed for by employing the method of partial correlation. As a rule, however, they are not easy to determine quantitatively, hence their dangerous character.

2. *Index Correlation**.

This is a form of spurious correlation which arises from the use of ratios for measurements. Thus if

$$z_1 = \frac{x_1}{x_3}, \quad z_2 = \frac{x_2}{x_3},$$

and x_1, x_2, x_3 are uncorrelated with one another, it can be shown that the correlation between z_1 and z_2 is

$$r_{z_1z_2} = \frac{\left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}{\sqrt{\left(\frac{\sigma_{x_1}}{\bar{x}_1}\right)^2 + \left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}\sqrt{\left(\frac{\sigma_{x_2}}{\bar{x}_2}\right)^2 + \left(\frac{\sigma_{x_3}}{\bar{x}_3}\right)^2}},$$

a quantity which may be as large as 0 5.

As an illustration from psychology, we may mention the correlation of intelligence quotients and achievement ratios†. An I Q is mental age divided by chronological age, an A R is educational age divided by mental age. Two I Q 's (found perhaps by two tests, or in successive years) will show $r = 0\ 5$ even if the mental ages are quite uncorrelated with one another or with age An I Q. and an A R may show $r = -0{\cdot}5$, for here mental age is denominator of one and numerator of the other Partial correlation, making the common element constant, will eliminate spurious index correlation.

* Karl Pearson, "On a form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs," *Proc Roy Soc* 1897, LX p. 489. In a note following Pearson's paper Sir Francis Galton illustrates the occurrence of index correlation by a simple and illuminating example

† See Pintner and Thomson, *Journ. Educ. Psychol* 1924, XV p. 433.

(6) VARIATE DIFFERENCE CORRELATION, OR THE ELIMINATION OF SPURIOUS CORRELATION DUE TO POSITION IN SPACE AND TIME*

This is a method for determining the correlation of variations from the "instantaneous mean," by correlating corresponding differences between successive values. If two variates x and y are such that

$$x = \phi (t) + X,$$
$$y = f (t) + Y,$$

where X and Y are the parts of x and y independent of the time t, then it can be shown that on certain not unreasonable assumptions,

$$r_{{}_mD_x {}_mD_y} = r_{XY},$$

if m is large enough, where ${}_mD_x$ is the mth difference between successive values of x, and similarly ${}_mD_y$. For example

x	${}_1D_x$	${}_2D_x$	${}_3D_x$	${}_4D_x$
47				
	6			
53		−4		
	2		4	
55		0		−5
	2		−1	
57		−1		
	1			
58				

Clearly the number of corresponding cases to be correlated is reduced each time. The method is only valid, *inter alia,* when there are so many cases in the x column that this reduction is immaterial. The following numerical example will it is hoped serve to make clear any obscure points in the above necessarily short account. The correlations between differences are worked out until r remains steady for several successive differences.

Numerical illustration of the Variate Difference Correlation Method†

The numbers headed *Savings* and *Tobacco* respectively are "indices" from Professor Georgio Mortara's article in the *Giornale degli Economiste e Rivista di Statistica,* February 1914. As they stand, they include a continuous secular increase both of savings and of consumption of tobacco which has occurred in Italy during the period in question, and their correlation is ·984. But first differences only correlate to an extent ·766 as is shown in detail in the following working:

* Miss F. E. Cave, *Proc. Roy Soc* 1904, LXXIV. p. 407; R. H. Hooker, *Journ. Roy. Stat. Soc* 1905, LXVIII p 396, "Student," *Biometrika,* 1914—15, X p. 179; and Anderson, *Biometrika,* 1914—15, X p. 269, where probable errors are given.

† From Beatrice M Cave and Karl Pearson, *Biometrika,* 1914—15, X. p 340.

Year	Savings x	${}_1D_x$	From 5	Sq	Product	Sq	From 3	${}_1D_y$	Tobacco y
1885	47								82
		6	1	1	1	1	1	4	
1886	53								86
		2	−3	9	3	1	−1	2	
1887	55								88
		2	−3	9	15	25	−5	−2	
1888	57								86
		1	−4	16	12	9	−3	0	
1889	58								86
		1	−4	16	8	4	−2	1	
1890	59								87
		1	−4	16	8	4	−2	1	
1891	60								88
		4	−1	1	2	4	−2	1	
1892	64								89
		1	−4	16	8	4	−2	1	
1893	65								90
		0	−5	25	20	16	−4	−1	
1894	65								89
		3	−2	4	8	16	−4	−1	
1895	68								88
		2	−3	9	9	9	−3	0	
1896	70								88
		3	−2	4	8	16	−4	−1	
1897	73								87
		2	−3	9	3	1	−1	2	
1898	75								89
		4	−1	1	1	1	−1	2	
1899	79								91
		4	−1	1	2	4	−2	1	
1900	83								92
		4	−1	1	0	0	0	3	
1901	87								95
		5	0	0	0	1	−1	2	
1902	92								97
		7	2	4	−2	1	−1	2	
1903	99								99
		8	3	9	0	0	0	3	
1904	107								102
		8	3	9	3	1	1	4	
1905	115								106
		12	7	49	0	0	0	3	
1906	127								109
		17	12	144	36	9	3	6	
1907	144								115
		11	6	36	36	36	6	9	
1908	155								124
		13	8	64	32	16	4	7	
1909	168								131
		12	7	49	28	16	4	7	
1910	180								138
		7	2	4	6	9	3	6	
1911	187								144
		5	0	0	0	16	4	7	
1912	192								151
		145		506	247	220		69	

In the table the variate x is the Savings Index, y is the Tobacco Index. We then have

$$\text{Mean of } {}_1D_x = 145/27 = 5{\cdot}37,$$
$$\text{Mean of } {}_1D_y = \ \ 69/27 = 2{\cdot}56.$$

Take 5 and 3 as provisional centres, so that $d_1 = {\cdot}37$, $d_2 = -{\cdot}44$. From deviations from these points we get, as shown in the table,

$$S({}_1D_x{}^2) = 506,$$
$$S({}_1D_x\,{}_1D_y) = 247,$$
$$S({}_1D_y{}^2) = 220.$$

Therefore the correlation (correcting for d_1 and d_2) is

$$r = \frac{247 + 5}{\sqrt{(506 - 4)}\sqrt{(220 - 5)}} = {\cdot}77.$$

Second differences have a small *negative* correlation, which increases till with *sixth* differences we reach $-{\cdot}431$, which seems to indicate that, when time has been eliminated, expenditure on tobacco in any year means *less* money saved.

Note, **1924.** *The coefficient of alienation* Kelley has in recent years emphasised the coefficient $k = \sqrt{(1 - r^2)}$, one effect of which is to inculcate caution in employing correlations. Consider the correlation r between two intelligence tests Till lately, a correlation of 0 8 would have been considered very satisfactory

If we know an individual's score on the first test, we can predict his most probable score on the second test But the actual performances of a number of such individuals in the second test will be scattered with a sigma equal to $k\sigma$ (pp 109 and 111 hereof), where σ is the sigma of all second test scores The probable error of our prediction, that is, is reduced to the fraction k of what it would be did we guess at random

Now for $r = 0{\cdot}8$, $k = 0{\cdot}6$, and this reduction to 0 6 of the error is not a very startling improvement on guessing. So $r = 0{\cdot}8$ means less to us, as an indication of predictive power, than we were accustomed to think It takes a correlation of 0 98 to reduce the probable error of a prediction to one-fifth of the probable error of a mere guess.

CHAPTER VIII

THE CORRECTION OF RAW CORRELATION COEFFICIENTS

Historical account—The elimination of irrelevant factors—Correction for observational errors (attenuation)—Correlation of gains and initial values.

(1) HISTORICAL ACCOUNT

THE history of the use of the theory of correlation in Psychology can hardly be said to have begun earlier than the commencement of the present century. During the previous twenty years, indeed, a great deal of work had been done by many observers in measuring simple mental abilities (by the "mental test" method) in larger or smaller groups of subjects, and attempts had even been made to determine in what way these abilities were related to one another and to more general mental ability, or "general intelligence."

Owing, however, to a universal lack of knowledge of the mathematical theory of correlation among psychologists during this period, the results were not obtained in a form suitable for comparison with one another, so that it is not surprising to find that they hopelessly contradict one another. The heterogeneity of the material worked with, the non-elimination of irrelevant factors, and the absence of any measure of the "probable error" of the results make the conclusions drawn by the investigators themselves from their researches utterly unreliable.

The first investigation showing any mathematical precision was that published by Clark Wissler* in 1901. It contained (*inter alia*) an account of the careful application of a large number of simple mental tests upon over 200 college students, and a correlation of the results with one another and with the students' marks in the various subjects of the college curriculum. The mental tests were found to correlate but slightly with one another or with ability in college subjects of study, though these latter showed considerable correlation with one another (·30—·75).

In the following year Aikens and Thorndike† published results which

* Clark Wissler, "The Correlation of Mental and Physical Tests," *Psychological Review, Monograph Supplement*, III No. 16, June 1901

† "Correlations among Perceptive and Associative Processes," *Psychological Review*, IX.

were in a sense confirmatory of those of Wissler, since, notwithstanding the greater similarity to one another of the functions investigated than in Wissler's research, the correlations were again found to be *low* For example, different tests devised for the measurement of "speed of association" were found to show hardly any correlation—a result which seemed to furnish some justification for the author's statement that "quickness of association as an ability determining the speed of all one's associations is a myth" (*op cit.* p. 375). A similar lack of close relationship was found in the case of other mental functions which would, on the evidence of introspection alone, be confidently classed as particular instances of the same general mental function.

In 1904 there appeared an epoch-making article by Professor C. Spearman*, the ideas originating in which have, at least in England, ever since dominated correlational work in its applications to psychology. Since we shall have frequent occasion in the course of this and the succeeding chapter to take exception to Professor Spearman's theories and mathematical methods, which appear to us incorrect and harmful, we may perhaps be allowed at this point, before embarking on controversial matters, to express our opinion that only Professor Spearman's enthusiasm and originality could have given to psychological correlation research the life and activity which it has shown during the last fifteen years. His work has stirred up both disciples and opponents to investigations which would otherwise never have occurred to them.

The new ideas in question fall into two main groups.

(1) *Corrections to the raw values of correlation coefficients.* Instead of measuring large numbers of individuals, as his predecessors had done, Professor Spearman contented himself with small numbers, groups of less than 40 in his first research, and as few as 11 in the second, but he proposes to make up for the unreliability thus introduced into the results by a more careful measurement of his cases, and the application of "corrections" to the "raw" values of his correlation coefficients by means of appropriate mathematical formulae.

(2) *The discovery of "hierarchical" order among correlation coefficients, and the* Theory of General Ability, *or the* Theory of Two Factors, *which has been built up on this foundation.* This theory has been a great incentive to research, and may possibly correspond to the facts, though we do not incline to think so. But its deduction from the occurrence of "hierarchical" order among the correlation coefficients is invalid, as

* "General Intelligence objectively determined and measured," *Amer. Journ. of Psychol* xv pp. 201—292.

will be shown in the next chapter In the present chapter we turn to the closer consideration of the first group of ideas, the correction of raw correlation coefficients

(2) THE ELIMINATION OF IRRELEVANT FACTORS

The first kind of correction is that for the elimination of irrelevant factors, and is nothing new in the theory of correlation, being simply the method of partial correlation described in the last chapter

For example, to eliminate the effects of difference of *age* in the group experimented upon, one would determine the partial correlation between the two characters under consideration, for "age constant," by means of the Yulean formula given on p. 137 A similar procedure is needed for eliminating the effects of difference of sex, etc A preferable course, however, would be to dispense as far as possible with the necessity for such corrections by selecting groups of individuals of the same age, sex, etc. Indeed, as Professor Spearman very properly points out*, the partial correlation formula must on no account be used in cases where there is too violent heterogeneity of the irrelevant factor. For example, we might with justice use it to eliminate the effects of age in a group where the extreme differences of age were only over a range of two or three years, but not in a group where the subjects ranged from say five years to fifteen years of age.

The second kind of correction introduced by Professor Spearman is a correction for what he calls "accidental" errors This correction is based on the "reliability coefficients" introduced by Professor Spearman

(3) CORRECTION FOR OBSERVATIONAL ERRORS (ATTENUATION)

It is by the aid of these reliability coefficients, as we have said, that Professor Spearman carries out the calculations which have for their object the elimination of observational errors.

It is clear that if the correlation between two series of quantities is really perfect, it must be reduced by observational errors and if imperfect, it will still tend to be reduced The amount of this reduction will probably be greater, the greater are the observational errors The size of these is however indicated to some extent by the reliability coefficients, so that it is but a short step to use these to correct for the reduction, and enlarge the correlation coefficient to its true value. An example given by Professor Spearman himself† shows this so clearly that it is not out of place to repeat it here

* "Demonstration of Formulae for the True Measurement of Correlation," *Amer Journ. of Psychol.* 1907, XVIII especially p 166.

† *Amer. Journ. of Psychol.* 1904, XV. p 271.

"A target was constructed of a great many horizontal bands, numbered from top to bottom. Then a man shot successively at a particular series of numbers in a particular order Clearly, the better the shot, the less numerical difference between any number hit and that aimed at, now, just as the measurement of any object is quite appropriately termed a 'shot' at its real value, so, conversely, we may perfectly well consider the series of numbers actually hit in the light of a series of measurements of the numbers aimed at. When the same man again fired at the same series, he thereby obtained a new and independent series of measurements of the same objects. Next, a woman had the same number of shots at some set numbers in a similar manner If, then, our above reasoning and formulas (see below) are correct, it should be possible, by observing the numbers hit and working out their correlations, to ascertain the exact resemblance between the series aimed at by the man and the woman respectively In actual fact, the series of numbers hit by the man turned out to correlate with those hit by the woman to the extent of 0·52; but it was noticed that the man's sets correlated with one another to 0·74, and the woman's sets with one another to 0·36, hence the true correspondence between the set aimed at by the man and that aimed at by the woman was not the raw 0·52, but

$$r = \frac{0{\cdot}52}{\sqrt{(0{\cdot}74 \times 0{\cdot}36)}} = 1{\cdot}00,$$

that is to say, the two persons had fired at exactly the same series of bands, which was really the case "

It will be seen that the formula for correction is "divide the raw correlation by the geometrical mean of the two reliability coefficients." This formula, or rather the more accurate formula of which it is an approximate form, we shall prove presently. Meanwhile it must be pointed out that only by a coincidence does the above corrected value happen to agree exactly with the true value. One of us (Thomson) has carried out a similar experiment a sufficiently large number of times to enable the distribution of the corrected coefficients to be compared with that of the raw values The "errors" were introduced by drawing cards. The "population" (corresponding to the number of shots) was throughout 32. The results were as shown in the table on the next page.

It will be seen that the corrected values are scattered considerably about the true value, even when the errors are as here uncorrelated, and that in many cases the corrected value is worse than the raw, though the median corrected value is better than the median raw value.

	Large errors		Small errors		Large errors		Small errors	
	Raw	Corrected	Raw	Corrected	Raw*	Corrected	Raw*	Corrected
Highest	0 656	1 34	0 678	0 88	0 746	1 37	0 827	1 19
Upper quartile	0 550	0 87	0 555	0 71	0 623	1 15	0 772	1 04
Median	**0 433**	**0 79**	**0 467**	**0 65**	**0 559**	**1 01**	**0 732**	**1 01**
Lower quartile	0 328	0 67	0 402	0 60	0 493	0 85	0 667	0 97
Lowest	0 163	0 44	0 281	0 46	0 327	0 68	0 579	0 82
No of values	25	100†	25	100†	20	30†	20	30†
True value	**0 667**	**0 667**	**0 667**	**0 667**	**1 000**	**1 000**	**1 000**	**1 000**

* These were also the reliability coefficients of columns 1 and 3

† There are more corrected than raw values because the latter can be grouped in various ways

The fact is, that the assumptions underlying Professor Spearman's formula may not be fulfilled in practice What these assumptions are can best be seen from the elegant proof given for the formula by Udny Yule*.

x_1 and y_1 are measures of x and y at a certain series of measurements, x_2 and y_2 are measures of x and y at another series of measurements.

Let
$$x_1 = x + \delta_1, \quad x_2 = x + \delta_2,$$
$$y_1 = y + \epsilon_1, \quad y_2 = y + \epsilon_2,$$

all terms denoting deviations from means.

Then, if it is assumed that δ, ϵ, the errors of measurement, are uncorrelated with one another or with x or y,

$$S(x\delta) \text{ etc.} = 0, \quad S(x_1 y_1) = S(xy).$$

Hence
$$r_{x_1y_1}\sigma_{x_1}\sigma_{y_1} = r_{xy}\sigma_x\sigma_y,$$
and similarly
$$r_{x_2y_2}\sigma_{x_2}\sigma_{y_2} = r_{xy}\sigma_x\sigma_y,$$
$$r_{x_1y_2}\sigma_{x_1}\sigma_{y_2} = r_{xy}\sigma_x\sigma_y,$$
$$r_{x_2y_1}\sigma_{x_2}\sigma_{y_1} = r_{xy}\sigma_x\sigma_y,$$
or
$$r_{xy}^4 = r_{x_1y_1}r_{x_2y_2}r_{x_1y_2}r_{x_2y_1}\frac{\sigma_{x_1}^2\sigma_{x_2}^2\sigma_{y_1}^2\sigma_{y_2}^2}{\sigma_x^4\sigma_y^4}.$$

But also, since $S(x\delta) = 0, \quad S(x_1x_2) = S(x^2)$,

and
$$r_{x_1x_2}\sigma_{x_1}\sigma_{x_2} = \sigma_x^2 \dagger$$
or
$$\sigma_{x_1}\sigma_{x_2} = \frac{\sigma_x^2}{r_{x_1x_2}} \text{ and } \sigma_{y_1}\sigma_{y_2} = \frac{\sigma_y^2}{r_{y_1y_2}},$$

* See Appendix (*e*) to Professor Spearman's article, *Brit Journ of Psychol* 1910, III p 294 There is a printer's mistake of omission in the last formula See also W Brown, "Some Experimental Results in Correlation," *Comptes Rendus du VI^me^ Congrès International de Psychologie*, Genève, 1910, where the same proof is quoted

† Whence Kelley's formula that true sigma is observed sigma times root of reliability, *J. of Educ. Psychol.* 1919, x p 229.

or
$$r_{xy}{}^4 = \frac{r_{x_1y_1}r_{x_2y_2}r_{x_1y_2}r_{x_2y_1}}{r^2_{x_1x_2}r^2_{y_1y_2}},$$

$$r_{xy} = \frac{\text{Geom Mean of correlation coefficients}}{\text{Geom. Mean of reliability coefficients}}.$$

It is this formula which is employed in Thomson's example given above. In Spearman's example, the numerator consisted only of $r_{(x_1+x_2)(y_1+y_2)}$, the subcalculations for $r_{x_1y_1}$ etc not being made

Attention must be drawn to the assumption that the errors of measurement δ and ϵ are uncorrelated with each other or with x or y

Now, these are very large assumptions to make Even in cases where the quantities δ, ϵ are genuine errors of measurement, there are strong reasons for assuming (on general principles and also from experimental evidence)* that they *will* be correlated. But in the case of almost all the simpler mental tests the quantities δ and ϵ are not errors of measurement at all They are the deviations of the particular performances from the hypothetical average performance of the several individuals under consideration Thus they represent the *variability* of performance of function *within* the individual. When an individual in the course of three minutes succeeds in striking through 100 *e*'s and *r*'s in a page of print on one day, and 94 under the same conditions a fortnight later, there is no error of observation involved The numbers 100 and 94 are the actual true measures of ability on the two occasions. The average or mean ability, which is the more interesting measure for the purposes of correlation, is doubtless different from either, but that does not make the other two measures erroneous. Evidently in these cases δ and ϵ represent *individual variability*, and to assume them uncorrelated with one another or with the mean values of the functions is to indulge in somewhat *a priori* reasoning.

There are two comparatively simple ways of testing the assumption:

$$(1) \quad S(x_1y_1) = S(xy) = S(x_2y_2),$$

$\therefore$ $S(x_1y_1) - S(x_2y_2)$ should $= 0$ within the limits of the probable error of the difference.

The probable error of $S(xy)$ is equal to

$$\cdot 67449 \sqrt{\left\{S(xy)^2 - \frac{S^2(xy)}{n}\right\}}.$$

In applying this to determine the probable error of the difference of

* See Karl Pearson, "On the Mathematical Theory of Errors of Judgment, with special reference to the Personal Equation," *Phil Trans* CXCVIII. A, pp. 235—299.

$S(x_1y_1)$ and $S(x_2y_2)$ one must bear in mind the possibility of these quantities being correlated.

$$(2)\quad r_{\substack{X_1-X_2\\Y_1-Y_2}} = \frac{S\{(x_1-x_2)(y_1-y_2)\}}{\sqrt{S(x_1-x_2)^2\ S(y_1-y_2)^2}}$$
$$= \frac{S\{(\delta_1-\delta_2)(\epsilon_1-\epsilon_2)\}}{\sqrt{S(\delta_1-\delta_2)^2 . S(\epsilon_1-\epsilon_2)^2}}$$
$$= 0,$$

if errors are uncorrelated with one another (since numerator then $= 0$).

Applying this test to a case of bisection and trisection, Brown gets

$$r_{\substack{B_1-B_2\\T_1-T_2}} = 0{\cdot}30 \pm 0{\cdot}09,$$

which proves the inapplicability of the formula here

Brown applied test (2) also to a case of correlation between speed of addition of figures and accuracy of addition in a group of 38 school-children (girls between the ages of 11 and 12) and found

$$r_{\substack{S_1-S_2\\A_1-A_2}} = 0{\cdot}35 \pm 0{\cdot}09.$$

Even when test (2) does give the value 0, we can only conclude from this that $\quad S(\delta_1\epsilon_1) + S(\delta_2\epsilon_2) = S(\delta_1\epsilon_2) + S(\delta_2\epsilon_1)$.

Replying to these and other criticisms of his formula for eliminating observational errors, Professor Spearman* admits that many forms of error will be correlated with each other and with the true values of the quantities measured. But these, he says, are generally of a continuously progressive nature, and he proposes to eliminate their influence by making at least three measurements of x, and taking the first and third of these together as the x_1 of the formula, the middle one as x_2, or to use more complicated but essentially similar devices No doubt, of course, this is a wise precaution to take, even though sceptics may still doubt whether the remaining so-called "accidental" variations are even yet uncorrelated with each other and with x and y

Brown†, using experimental data obtained by two independent observers estimating the lengths of lines, found a considerable correlation ratio between the errors of observation and the true lengths of the lines, the regression being very far from linear He is of opinion that "no assumptions as to the correlation or non-correlation of such deviations are in the least justified."

The correlation ratios between lengths and errors, and the corresponding regression curves, are shown on the opposite page‡.

In a review of Brown's article Mr J. R. Wilton§ has made some

* *Brit. Journ Psychol* 1910, III. p. 271 † *Ibid.* 1913, VI. p 223.

‡ From *Brit. Journ Psychol* 1913, VI. pp 236—8.

§ *Journ. of Exp Pedag.* 1914, II. p 302.

ingenious suggestions for weighting the different x's in a manner suggested by quadrature formulae He remarks also that a grouping should be sought which would as far as possible satisfy the assumptions.

It may finally be pointed out that *even if the errors δ and ϵ were known with certainty to be uncorrelated with each other and with the true values, yet, with such small numbers of cases as are used in many of the psychological researches in which Professor Spearman's formula has been employed, the chance of, e g $S(x_1y_1)$ being nearly equal to $S(x_2y_2)$ is exceedingly small*, and it is difficult to attach any meaning to an *artificial, post hoc* separation of the data into halves such that this condition is satisfied (and also others). The formula is at any rate inapplicable to samples such as 30, 24, 52, which have been freely used in experimental work, or if used at all, it can only be as a guide to the sufficiency of the sample. It is an essential of good work to use such samples that corrections to the raw values obtained are unimportant.

Regression Curves of Correlation Table

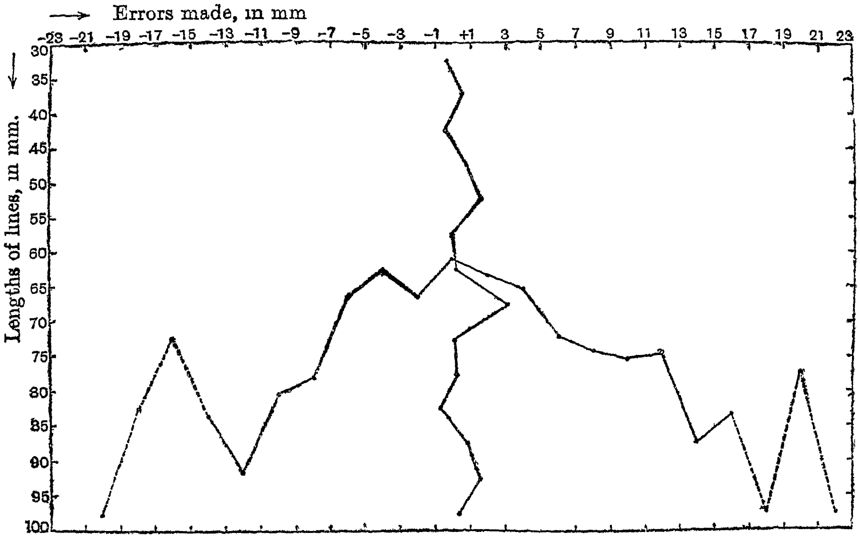


Fig 27

$r = 0\ 033 \pm 0\ 030,$
$\eta_x = 0\ 182 \pm 0{\cdot}029,$
$\eta_y = 0{\cdot}323 \pm 0{\cdot}027.$

(4) AN ARTIFICIAL ILLUSTRATIVE EXAMPLE

Many of the above considerations may be illustrated by the following artificial example.

First let us remind the reader that following the binomial law discussed in Chapter II, *symmetrical* distributions of one variable run in such ways as 1 · 2 · 1 if there are three values, or 1 3 · 3 1 if there are four. If x and y are uncorrelated, and 16 cases of x are distributed over three values thus

4	8	4

then each of these will be distributed vertically as regards y in similar fashion giving

	$- x \rightarrow$		
$\mid$	1	2	1
y	2	4	2
$\downarrow$	1	2	1

which is a table showing zero correlation of x and y.

(*a*) *Errors uncorrelated Attenuation correction permissible.*

Consider now a correlation table for x and y for which true $r = 0{\cdot}667$ thus:

16	16	0	0
16	48	32	0
0	32	48	16
0	0	16	16

Taking the columns and rows as unit distance apart we have

$$r = \frac{S\,(xy)}{\sqrt{S\,(x^2)\,S\,(y^2)}} = \frac{128}{192} = \frac{2}{3}.$$

Suppose that *uncorrelated* errors splash each 16 cases over an area thus·

1	2	1
2	4	2
1	2	1

where the unit is the same The resulting observed correlation table is as follows and gives $r = 0{\cdot}4$:

1	3	3	1	0	0
3	11	15	9	2	0
3	15	28	24	9	1
1	9	24	28	15	3
0	2	9	15	11	3
0	0	1	3	3	1

Again taking the columns and rows as unit distance apart we have

$$r = \frac{S\,(xy)}{\sqrt{Sx^2 Sy^2}} = \frac{128}{320} = 0{\cdot}4.$$

Notice that $S\,(xy)$ is unaltered by the attenuation. It is $S\,(x^2)$ and $S\,(y^2)$ which change, i e. σ changes

It can by similar methods be shown that the reliabilities are $r_x = r_y = 0{\cdot}6$

For were there no errors, the reliability would be unity and the correlation of either variable with its own identical and correctly measured self would give a correlation table like this:

32	0	0	0
0	96	0	0
0	0	96	0
0	0	0	32

Errors of the same nature as those above would splash each 32 or 96 cases over an area and the *observed* table would be

2	4	2	0	0	0
4	14	16	6	0	0
2	16	32	24	6	0
0	6	24	32	16	2
0	0	6	16	14	4
0	0	0	2	4	2

and this table gives $r_x = 0\ 6$.

By Spearman's formula the corrected value for r_{xy} is then $0{\cdot}4/0{\cdot}6 = 0{\cdot}667$, which agrees with the true value.

(*b*) *Errors correlated. Attenuation correction misleading.*

Consider now, however, this parallel case in which the true correlation of x and y is only 0·33 thus:

8	16	8	0
16	40	32	8
8	32	40	16
0	8	16	8

Let *correlated* errors splash each 8 cases over an area thus:

1	1	0
1	2	1
0	1	1

If the reader will carry this out he will obtain this correlation table:

1	3	3	1	0	0
3	11	15	9	2	0
3	15	28	24	9	1
1	9	24	28	15	3
0	2	9	15	11	3
0	0	1	3	3	1

identical with that obtained under case (*a*) and again giving $r = 0{\cdot}4$.

The reliabilities however are now different and can be seen to be $r_x = r_y = 0{\cdot}8$.

For the *true* reliability table

32	0	0	0
0	96	0	0
0	0	96	0
0	0	0	32

is converted by these *correlated* errors into

4	4	0	0	0	0
4	20	16	0	0	0
0	16	40	24	0	0
0	0	24	40	16	0
0	0	0	16	20	4
0	0	0	0	4	4

giving r_x or $r_y = 0{\cdot}8$.

If we correct the observed r_{xy} for attenuation therefore we obtain $r = 0{\cdot}4/0{\cdot}8 = 0{\cdot}5$, whereas the true value is 0·33 and the correction is in the wrong direction. This undesirable phenomenon would seem to be

particularly liable to occur when attenuation corrections are made by splitting a test, since a boy who is off colour in one part is very likely to be also off colour in the other and hence errors will be correlated

(5) CORRELATION OF INITIAL VALUE OF A MENTAL FUNCTION WITH ITS GAIN OVER A CERTAIN PERIOD

Observational errors, which if uncorrelated tend to deflect an ordinary correlation inward toward zero, further tend, in the special case of correlation of gains with initial values, to depress r towards -1, as was discovered by Thorndike* The two effects augment one another in the case of positive correlations but are opposed if the correlation is negative. The present correction† is particularly important in correlations concerning learning and practice, annual changes in intelligence, etc.

Let $a =$ true initial value,

$g =$ true gain,

$a + g =$ true final value,

$e =$ error of measurement in initial value,

$e' =$ error of measurement in final value,

$x =$ measured initial value,

$y =$ measured gain,

$z =$ measured final value,

then

$$z = x + y,$$
$$x = a + e,$$
$$z = a + g + e',$$
$$y = g + e' - e.$$

The reason for the reduction in the apparent correlation of x and y is the presence of $+e$ in x and of $-e$ in y. If a and g are really uncorrelated, this will cause an apparent negative correlation

Thomson has then shown that, on the assumption that e and e' are not correlated with one another or with g or a (assumptions similar to those made in using Spearman's attenuation formula),

$$\text{true } r_{ag} = \frac{\sigma_y r_{xy} + \sigma_x (1 - r_x)}{\sqrt{r_x}\sqrt{\{\sigma_y^2 - \sigma_x^2 (1 - r_x) - \sigma_z^2 (1 - r_z)\}}}$$

and also

$$= \frac{\sigma_z r_{xz} - \sigma_x r_x}{\sqrt{r_x}\sqrt{(r_x\sigma_x^2 + r_z\sigma_z^2 - 2\sigma_x\sigma_z r_{xz})}},$$

where r_x and r_z are the reliability coefficients of x and z That is, r_x is the correlation of two measurements of x and r_z is the correlation of two measurements of z. r_{xz} must not be used as a reliability coefficient here.

* *Journ. Exp. Psychol* 1924

† Godfrey H. Thomson, *Journ. Exp Psychol* 1924, VII p 321.

CHAPTER IX

THE THEORY OF GENERAL ABILITY

This apparent unity is illusory Man, in fact, is a microcosm as complex as the world which is mirrored in his mind, he is a federation incompletely centralised, a hierarchy of numerous and conflicting passions, each of which has ends of its own, and each of which, separately considered, would give a different law of conduct He is in some sense a unit, but his unity is such as to include an indefinite number of partly independent sensibilities.

LESLIE STEPHEN, *The Science of Ethics*, p. 69

Discovery of "Hierarchical" order among correlation coefficients—Use of the formula for the correction of observational errors to prove the existence of a general factor—Researches between 1904 and 1912—A criterion for hierarchical order applied to numerous researches—Complications in the original theory

(1) DISCOVERY OF HIERARCHICAL ORDER AMONG CORRELATION COEFFICIENTS*

THE controversy as to whether ability in any individual is general, or specific, or in groups or "faculties" is a very old one, but for the purposes of the present chapter it is not necessary to go back prior to 1904, in which year there was published the first† of a series of articles in which Professor C. Spearman has developed his Theory of General Ability, or Theory of Two Factors, as it is alternatively named.

Professor Spearman's method in that paper was to measure a number of mental abilities, some of them school subjects, others artificial tests, in a number of persons, and calculate the correlation coefficients of each of these activities with each of the others. These correlation coefficients, he then noticed, had a certain relationship among themselves, a relationship which may be called hierarchical order, and is explained in detail later. He saw, quite rightly, that the presence of a general factor would produce this hierarchical order among the coefficients, and, reversing this argument, he concluded that the presence of hierarchical order proved the existence of a general factor.

In this first series of investigations Professor Spearman used the following groups of subjects: 24 village school-children of both sexes,

* Sections 1, 3 and 5 of this chapter are largely extracts from an article by G. H. Thomson in the *Psychological Review*, 1920, xxvii. p. 173.

† "General Intelligence objectively determined and measured," C. Spearman, *Amer. Journ. Psychol.* 1904, xv pp. 201—293.

age limits 10 0 to 13·10; 23 boys of a high class preparatory school, age limits 9·5—13·7, and 27 adults of both sexes, age limits 21—78 The tests employed were those for pitch discrimination, weight discrimination, and discrimination of light intensities, and measures of intelligence were obtained, in the case of the children, from results of school examinations, grading by teachers, and grading of one another by the children themselves (measure of common sense)

The various school subjects in the preparatory school were found to correlate highly with one another, and when, with the inclusion of pitch discrimination and music, they were arranged in rows and columns, it was found possible to place them in such an order that the correlation coefficients formed a *hierarchy*, each being (with very few exceptions) greater than any to the right of it in the same row, or below it in the same column, thus

	Classics	French	English	Math	Discrim	Music
Classics	—	0 83	0 78	0 70	0 66	0 63
French ...	0 83	—	0 67	0 67	0 65	0 57
English .	0 78	0 67	—	0 64	0 54	0 51
Math .. .	0 70	0 67	0 64	—	0 45	0 51
Discrim .	0 66	0 65	0 54	0 45	—	0 40
Music	0 63	0 57	0 51	0 51	0 40	—

This fact of "hierarchical" order which he had thus discovered was taken by Professor Spearman to indicate the presence of some common fundamental function which saturates in different degrees the different activities, and is the sole cause of correlation between them except in the case of very similar activities

It can easily be shown that if all the correlations are due solely to one common or general factor, then the correlation coefficients will be in perfect hierarchical order*

Let a and p be two mental tests or other activities, and g be the general factor. Then $r_{ap\,g}$ is the correlation that a would have with p for constant g, and equals

$$\frac{r_{ap} - r_{ag}r_{pg}}{\sqrt{(1 - r_{ag}^2)}\sqrt{(1 - r_{pg}^2)}}.$$

But if g is the sole source of correlation, $r_{ap\,g}$ must be zero, i.e.

$$r_{ap} = r_{ag} \cdot r_{pg}$$

Similarly $$r_{bp} = r_{bg}r_{pg}.$$

Hence $$\frac{r_{ag}}{r_{bg}} = \frac{r_{ap}}{r_{bp}} \text{ and similarly } = \frac{r_{aq}}{r_{bq}}.$$

* Hart and Spearman, *Brit. Journ. Psychol* 1912, v. p 58 quoting Yule. Previous though less satisfactory proofs had also been given by Spearman.

A little consideration of this last equation shows that, if it be true for *any* four of the tests, it implies the possibility of arranging the correlation coefficients in the order we have termed "hierarchical" and more than this, that the values of r in any one column of the "hierarchy" will bear a constant ratio, each to each, to their partners in any other column.

Since clearly *perfect* hierarchical order cannot be expected in any *experimental* research, it becomes important to know what deviation from perfection can be allowed without giving up the idea of a general factor· or on the other hand, what approach to perfection can be attained without the presence of a general factor. These questions will occupy us presently Meanwhile we turn for a while to another form of argument used by Professor Spearman.

(2) USE OF THE FORMULA FOR THE CORRECTION OF OBSERVATIONAL ERRORS TO PROVE THE EXISTENCE OF A GENERAL FACTOR

He considered that by the use of his formula for the correction of observational errors he could demonstrate the same thing (the existence of a "central function"), and could in particular show "that the common and essential element in the intelligences wholly coincides with the common and essential element in the sensory functions*" The method of proof is as follows

Let x_1, x_2 be two distinct measures of sensory discrimination, and y_1, y_2 two distinct measures of intelligence.

Then the correlation of the function common to the functions measured by x_1, x_2 with the function common to the functions measured by y_1, y_2 is equal to

$$\frac{\sqrt[4]{r_{x_1y_1}r_{x_1y_2}r_{x_2y_1}r_{x_2y_2}}}{\sqrt{r_{x_1x_2}r_{y_1y_2}}},$$

and if the two functions referred to are identical this expression should be equal to *unity*.

In the present article, Spearman uses a simplified formula,

$$\frac{r_{xy}}{\sqrt{r_{x_1x_2}r_{y_1y_2}}},$$

and puts for the numerator the *average* of the various correlations

* *Amer. Journ. Psychol.* 1904, xv p. 269

evaluated between the intelligences and the discriminations, and in the denominator puts

$r_{x_1x_2}$ = the average correlation of the intellective gradings with one another,

$r_{y_1y_2}$ = the average correlation of the gradings in discrimination with one another,

and in this way gets results approximately equal to 1 in the different groups tested.

One or two remarks may appropriately be made here In the first place, the full formula is the only one that can be used with any meaning or justice since it is the only one which issues logically from the mathematical proof. In the second place, the applicability of the true formula must be considered in the light of its presuppositions (mentioned above, p. 158).

Indeed, the assumption that δ and ϵ are uncorrelated with each other or with x or y seems even more unwarrantable here than in the case of "correcting" coefficients, for which the formula was originally devised

(3) RESEARCHES BETWEEN 1904 AND 1912

A number of experimental researches on these lines, in some of which Professor Spearman himself took part, were carried out during the eight years following 1904, but with very conflicting results, some experimenters finding the hierarchical order among the coefficients, others finding no such order Two articles of this period, for example, are those of Mr Cyril Burt*, who found practically perfect hierarchical order, and Dr William Brown†, who found small trace of such order A similar conflict of opinion was found with regard to the alternative method of attack, as for example in the research by Messrs Thorndike, Lay and Dean‡. The subjects examined were 37 young women students and 25 high school boys The tests for sensory discrimination were:

(1) drawing lines equal to given lines, and

(2) filling boxes with shot to equal in weight standard weights;

those for intelligence were:

(3) judgment of fellow-students, and

(4) judgment of teachers.

* Cyril Burt, "Experimental Tests of General Intelligence," *Brit. Journ Psychol* 1909, III. pp. 94—177.

† William Brown, "Some Experimental Results in the Correlation of Mental Abilities," *Brit. Journ. Psychol* 1910, III. pp. 296—322.

‡ Thorndike, Lay and Dean, "The Relation of Accuracy in Sensory Discrimination to General Intelligence," *Amer. Journ. Psychol.* July 1909, XX. pp. 364—369.

For the high school boys (3) and (4) were combined teachers' and fellow-students' judgments and school marks, respectively

In the first case, Spearman's formula gave for the correlation of the factor common to (1) and (2) with that common to (3) and (4) *the value* 0 26 *instead of* 1·00. In the second case, the value was 0 29 Moreover, Thorndike found a much higher correlation between discrimination of lengths and discrimination of weights than between either one of them and general intelligence, the coefficients being

Accuracy in drawing lines, intelligence	0·15,
Accuracy in making up weights, intelligence ...	0 25,
Accuracy in drawing lines and making up weights	0·50

Thus the results were in decided conflict with both parts of Spearman's concluding statement "that all branches of intellectual activity have in common one fundamental function (or group of functions) whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others*."

Thorndike sums up as follows: "In general there is evidence of a complex set of bonds between the psychological equivalents of both what we call the formal side of thought and what we call its content, so that one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatever* common to all mental functions, or to any part of them†."

Things were in this very unsatisfactory state when an important article by Professor Spearman, in cooperation with Dr Bernard Hart, appeared in 1912‡. In this article the difficulty of making an unbiassed judgment as to the presence or absence of hierarchical order was recognised, and a form of calculation was given for obtaining a numerical criterion of the degree of perfection of hierarchical order, which criterion would be independent of any bias on the part of the calculator and would, it was hoped, give the *true* amount of hierarchical order, corrected for the sampling errors of experiment. This criterion ranges theoretically from zero, for absence of hierarchical order, to unity, for perfection of hierarchical order. But their formula can, arithmetically, exceed unity.

* C Spearman, *Amer Journ Psychol* xv p 284 † *Op cit* p 368.

‡ "General Ability, its Existence and Nature," by B. Hart and C Spearman, *Brit. Journ. Psychol.* 1912, v pp 51—84

Note, 1924 From correspondence and conversation we gather that the criterion for hierarchical order discussed in the next few pages has now been abandoned in practice by Professor Spearman who prefers to employ the exact criterion given on p 165 at the foot, which he originally communicated to Burt for his paper in *Brit. Journ Psychol* 1910, III. p. 159 Though Spearman has not yet as far as we are aware published any survey of correlation coefficients by this method to replace his survey (with Hart) by the column correlation method, his paper with Holzinger on the probable error of the new criterion is an important step towards making such a survey possible. See p 192 *et seq.*

(4) A CRITERION FOR HIERARCHICAL ORDER

The underlying idea was that if the above square table[*] of correlation coefficients shows hierarchical order in any degree, there will be correlation between the columns of that table taken in pairs, and that when the hierarchical order is perfect the columnar correlation R will rise to unity, except in so far as it is blurred by the sampling errors, which obviously cannot increase an already perfect correlation, but can only decrease it Let us write dashed letters throughout for the true values of the various quantities, which in ordinary experiment are unknown, reserving undashed letters for their measured values. We then have.

r' = true correlation coefficient,

e = its sampling error on one occasion, so that

$r = r' + e$,

$\overline{r'}$ = mean of the column of true values r',

$\bar{r}$ = mean of the column of observed values r.

In finding these means, that coefficient is omitted which has no partner in the column with which correlation is being found. Write also

$\rho' = r'$ measured from the mean of the true column, i.e.

$= r' - \overline{r'}$, and similarly

$\rho = r$ measured from the mean of the observed column, i.e.

$= r - \bar{r}$,

$\epsilon = \rho - \rho', = e - \bar{e}$,

where $\bar{e}$ is the mean of the column of e's.

Then for two columns a and b, the true columnar correlation which we desire to know is

$$R_{ab}' = \frac{S\,(\rho_{xa}'\rho_{xb}')}{\sqrt{\{S\,(\rho_{xa}'^2)\,S\,(\rho_{xb}'^2)\}}} \qquad \ldots\ (1),$$

by the Bravais-Pearson product-moment formula, S indicating summation over the various values of x, i e. summation up the column. This can be written

$$R_{ab}' = \frac{S\,(\rho_{xa}\rho_{xb}) - S\,(\epsilon_{xa}\epsilon_{xb}) - S\,(\rho_{xb}'\epsilon_{xa}) - S\,(\rho_{xa}'\epsilon_{xb})}{\sqrt{\{S\,(\rho_{xa}\rho_{xa}) - S\,(\epsilon_{xa}\epsilon_{xa}) - 2S\,(\rho_{xa}'\epsilon_{xa})\}}\sqrt{\{S\,(\rho_{xb}\rho_{xb}) - S\,(\epsilon_{xb}\epsilon_{xb}) - 2S\,(\rho_{xb}'\epsilon_{xb})\}}}.$$

In this expression, the three quantities of the form $S\,(\rho\rho)$ are known. The three quantities of the form $S\,(\epsilon\epsilon)$ are not known, but an attempt can be made to estimate their probable values from the known standard deviations of the correlation coefficients. The four quantities of the

[*] p. 165.

form $S(\rho'\epsilon)$ are treated by Dr Hart and Professor Spearman, in their paper, as negligible, on the ground that ρ' will not in general be correlated with ϵ. This assumption we suggest was erroneous.

The formula at which Dr Hart and Professor Spearman eventually arrive, after neglecting these quantities and making various other assumptions, is

$$R_{ab}' = \frac{S(\rho_{xa}\rho_{xb}) - (n-1)\, r_{ab}\overline{\sigma_{xa}\sigma_{xb}}}{\sqrt{\{S(\rho_{xa}^2) - (n-1)\,\overline{\sigma_{xa}^2}\}}\sqrt{\{S(\rho_{xb}^2) - (n-1)\,\overline{\sigma_{xb}^2}\}}} \quad (2),$$

where the σ's are standard deviations of the correlation coefficients, the bar indicates mean values for the column, and n is the number of pairs of correlation coefficients concerned, in the two columns. In using their formula, its authors do not apply it to all the pairs of columns in the square table. They say: "In any case the correction must be kept within limits as usual, the larger the correction the less it is to be trusted. If the sampling errors are large enough, they eventually will quite swamp the true differences of magnitude upon which the observed correlation should be based. In this case, the true correlation is beyond ascertainment; any attempt at correction is merely illusory. To avoid this, and at the same time to ensure impartial treatment of all data, it is necessary to fix beforehand some definite limit to the feasibility of correction. We have here adopted the following standard. in order to attempt to estimate the correct correlation between columns, *it is required that in each of these columns the mean square deviation should be at least double the correction to be applied to that deviation.*"

That is to say, the equation (2) is not to be used unless, in each factor of the denominator, $S(\rho^2)$ is at least double its correction $(n-1)\,\overline{\sigma^2}$. This condition (the "correctional standard") will be found to be important.

The authors applied their criterion to all the experimental work available, work dating from various periods, and representing the researches of 14 experimenters on 1463 men, women, boys and girls. From beginning to end the values of the criterion were positive and very high. The mean was almost complete unity. That is to say, Dr Hart and Professor Spearman claimed that all the data then available showed perfect hierarchical order among the correlation coefficients, even the data of workers like Dr Brown and Professor Thorndike, who had been unable to detect any such order. The reasons why the hierarchical order among the correlation coefficients was not obvious at a glance were, according to these authors, two. In the first place, their theory did not

entirely deny the presence of Group Factors of narrow range, and tests which were too similar were, according to them, to be pooled, before the hierarchical order would become apparent Only in very few cases however did they find it necessary to pool tests in the data used In the second place, the obscuring of the perfect hierarchical order was, according to them, due to the fact that only a small sample of subjects is examined For this error allowance is made in the formula for calculating their criterion

Dr Hart and Professor Spearman therefore considered their "Theory of Two Factors" proved This theory considers ability in any activity to be due to two factors One of these is a General Factor, common to all performances The other is a Specific Factor, unique to that particular performance, or at any rate extending only over a very narrow range including only other very similar performances "It is not asserted," they say, "that the General Factor prevails exclusively in the case of performances too alike, but only that when this likeness is diminished, or when the resembling performances are pooled together, a point is soon reached where the correlations are still of considerable magnitude, but now indicate no common factor except the General one"

In the same paper Dr Hart and Professor Spearman consider, and in their opinion confute, two other theories, (*a*) the older view of Professor Thorndike, viz a general independence of all correlations, and (*b*) Professor Thorndike's newer view of "levels," or the almost universal belief in "types." If the former were true, their criterion would, they consider, show an average value of about zero if the latter, a low minus value.

Their argument runs as follows.

If none but quite Specific Factors are present, the correlations will all be zero, and the pairs of columns will show no correlation with one another. If however correlations exist, but are due to Group Factors alone, then tests which share a Group Factor will correlate highly, but others will not correlate at all Let there be three such Group Factors, then we shall obtain not a hierarchy but an arrangement like this:

	S_1	S_2	S_3	A_1	A_2	A_3	D_1	D_2	D_3
S_1	.	*h*	*h*	*l*	*l*	*l*	*l*	*l*	*l*
S_2	*h*	.	*h*	*l*	*l*	*l*	*l*	*l*	*l*
S_3	*h*	*h*	.	*l*	*l*	*l*	*l*	*l*	*l*
A_1	*l*	*l*	*l*	.	*h*	*h*	*l*	*l*	*l*
A_2	*l*	*l*	*l*	*h*	.	*h*	*l*	*l*	*l*
A_3	*l*	*l*	*l*	*h*	*h*	.	*l*	*l*	*l*
D_1	*l*	*l*	*l*	*l*	*l*	*l*	.	*h*	*h*
D_2	*l*	*l*	*l*	*l*	*l*	*l*	*h*	.	*h*
D_3	*l*	*l*	*l*	*l*	*l*	*l*	*h*	*h*	.

h=high correlation. *l*=low correlation. See *Brit. Journ Psychol.* 1912, v. p. 57.

in which the high correlations are concentrated along the diagonal. In this arrangement some columns will correlate positively, namely those in which the high correlations come opposite one another, but these will be in the minority and most pairs of columns will correlate negatively. Professor Spearman and Dr Hart conclude therefore that in the absence of a General Factor the average correlation between columns will be either zero or negative, and that only a General Factor will give a very high positive correlation between pairs of columns.

In this consideration of Group Factors however it has been tacitly assumed that there is no overlapping of such factors. If this were so then indeed a hierarchy would be impossible But it is at any rate a conceivable hypothesis that such overlapping should occur, that for example there might exist a factor common to three tests *a*, *b*, *c* and another common to *c*, *d*, *e*, so that *c* contains both factors and on this hypothesis an excellent hierarchy can be obtained without any General Factor, and the average column correlation can even approach unity, as we shall show presently.

(5) COMPLICATIONS IN THE ORIGINAL THEORY

Many experimental researches were inspired by this paper of Dr Hart and Professor Spearman, of which, as a good example, may be cited one in 1913 by Mr Stanley Wyatt*. It is not too much to say that in practically all of these the application of the Hart and Spearman criterion gave values closely approximating to unity and therefore supporting the Theory of General Ability. But complications began to arise, of which the first of importance will be found in Dr Edward Webb's monograph on "Character and Intelligence," in 1915† Dr Webb considered that he had found (in addition to Professor Spearman's General Ability) a second general factor, which he calls "persistence of motives." Other writers began to find that their data required for their explanation large Group Factors, of wider range than those contemplated in the original form of Professor Spearman's theory‡. Quite recently Mr J. C. Maxwell Garnett, discussing the data of a number of workers with the aid of mathematical devices which he has introduced for the purpose,

* Stanley Wyatt, "The Quantitative Investigation of Higher Mental Processes," *Brit. Journ Psychol* 1913, VI pp 109—133.

† E Webb, "Character and Intelligence," *Brit Journ. Psychol*, *Monog. Supplement*, 1915, No. 3, pp. IX and 99

‡ See especially N Carey, "Factors in the Mental Processes of School Children," *Brit. Journ. Psychol* 1916, VIII pp 170—182.

concludes that in addition to the single general factor of Professor Spearman, there are two large Group Factors which are practically general* (one of them being indeed almost identical with Dr Webb's second general factor), which he calls respectively "Cleverness" and "Purpose," both distinct from General Ability.

It is clear therefore that in any case the simple original form of Professor Spearman's theory is becoming complicated by additions which tend to modify it very considerably. Meanwhile, however, one of us had come definitely to the conclusion that the mathematical foundations upon which it was based were in fact incorrect. Before developing the line of argument which led to this, it will be well to re-state Professor Spearman's case in its simplest terms in a few words

It is entirely based upon the observation and measurement of hierarchical order among correlation coefficients It states that after allowance has been made for sampling errors this hierarchical order is found practically in perfection. And it finally states that such a high degree of perfection can only be produced by a General Factor, and the absence of Group Factors, which would mar the perfection.

* J. C Maxwell Garnett, "General Ability, Cleverness, and Purpose," *Brit Journ Psychol.* 1919, IX pp 345—366.

Note, 1924 There is not space, in a reprint in which only minor alterations are possible, to refer to several points which have since been raised concerning the theories discussed in this and the following chapter But we may perhaps be permitted to say three things concerning a paper of Dr Spearman's in *Brit Journ Psychol* 1922, XIII p 26 In the first place, that paper gives a proof of what Thomson has always said, that a general factor is a possible but not a necessary explanation of the correlational facts And secondly, Dr Spearman does not perhaps represent Thomson's position completely, as we think anyone will agree who first reads Spearman's paragraph at the top of his page 30, and then reads pages 175, 176, 177 of this book, published a year before he wrote, and indeed originally in *Proc Roy Soc* 1919 The point of agreement which Dr Spearman stresses in his last paragraph however we are glad to welcome.

CHAPTER X*

A SAMPLING THEORY OF ABILITY

The case against the validity of Professor Spearman's argument—Hierarchical order produced by random overlap of group factors, without any general factor—Application of the "criterion" to these cases, apparently proving the presence of a general factor—The erroneous nature of the "criterion"—Hierarchical order the natural order among correlation coefficients—A sampling theory of ability—Transfer of training—Conclusions

(1) THE CASE AGAINST THE VALIDITY OF PROFESSOR SPEARMAN'S ARGUMENT

As we have already seen in previous chapters, it is possible, by means of dice throws or in other ways, to make artificial experiments on correlation, with the immense advantage that the machinery producing the correlation is known, and that therefore conclusions based upon the correlation coefficients can be confronted with the facts. Working on these lines, one of us† made, in 1914, a set of imitation "mental tests" (really dice throws of a complicated kind), which were known to contain no General Factor. The correlations were produced by a number of Group Factors which were of wide range, and, unlike Professor Spearman's Specific or Narrow Group Factors, they were not mutually exclusive.

These imitation mental tests, containing no General Factor, gave however a set of correlation coefficients in excellent hierarchical order, and the criterion was when calculated found to be unity, so that had these correlation coefficients been published as the result of experimental work, they would have been claimed by Professor Spearman as proving the presence of a General Factor. In a short reply Professor Spearman laid stress on the fact that this arrangement of Group Factors which thus produced practically perfect hierarchical order was not a random arrangement, that it was exceedingly improbable that this one special arrangement should have occurred in each of the psychological researches of many experimenters, so improbable indeed as to be ruled entirely out of court‡, and that a random arrangement of Group Factors, though

* Much of this chapter, and part of the preceding, consists of extracts from "General versus Group Factors in Mental Activities," by G. H Thomson, *Psychol. Review*, 1920, XXVII. p. 173.

† Godfrey H. Thomson, "A Hierarchy without a General Factor," *Brit. Journ. Psychol* 1916, VIII pp. 271—281.

‡ C Spearman, "Some Comments on Mr Thomson's Paper," *Brit. Journ. Psychol.* 1916, VIII. p. 282.

it might give some hierarchical order, would not give it in the perfection actually found The obvious way to find out if this is so or not is *to try it*, with artificial "mental tests" formed of dice throws This was done in November and December of 1918, after an unavoidable delay of some years Sets of artificial variables (analogous to the scores in mental tests) were made, in each of which the arrangement of Group Factors was decided by the chance draws of cards from a pack*. It was found that hierarchical order resulted, which when measured by the "criterion" appeared to be perfect.

(2) HIERARCHICAL ORDER PRODUCED BY RANDOM OVERLAP OF GROUP FACTORS, WITHOUT ANY GENERAL FACTOR†

Write down the letters x_1, x_2, x_3, as the names of the variates to be formed, and prepare columns to receive the numbers of group factors and specific factors in each variate Determine each number by the draw of a card from an ordinary playing pack, returning the card and shuffling between each draw. the knave, queen, and king may be counted 11, 12, and 13 respectively. The result of one such set of drawings is shown in this table:

Variate	Group factors	Specific factors	Total
x_1	5	5	10
x_2	5	3	8
x_3	12	12	24
x_4	1	3	4
x_5	7	6	13
x_6	9	5	14
x_7	13	13	26
x_8	1	2	3
x_9	9	3	12
x_{10}	11	5	16

Proceed next to identify the group factors of each variate. Do this by using a single suit of the pack. After shuffling it well, lay out the top five cards to represent the five group factors in x_1, and note them.

* Godfrey H. Thomson, "On the Cause of Hierarchical Order among the Correlation Coefficients of a Number of Variates taken in Pairs," *Proceedings of the Royal Society of London*, 1919, xcv. A, pp 400—408. See also, by the same author, "The Hierarchy of Abilities," and "The Proof or Disproof of the Existence of General Ability," in *Brit. Journ. Psychol.* 1919, IX pp. 321—344

† This section and also section 5 consists largely of extracts from the *Proc. Roy. Soc.* 1919, xcv. A, pp. 400—408.

After replacing them and reshuffling, do the same for x_2, and so on, as in this table:

	Ace	2	3	4	5	6	7	8	9	10	Kn	Q	K
x_1	/	/			/		/						/
x_2					/		/		/	/		/	
x_3	/	/	/	/	/		/	/	/	/	/	/	/
x_4												/	
x_5	/	/			/		/	/				/	/
x_6				/	/	/	/	/	/	/	/	/	
x_7	/	/	/	/	/	/	/	/	/	/	/	/	/
x_8		/											
x_9	/	/	/	/	/	/	/	/					/
x_{10}		/	/	/	/	/	/	/		/	/	/	/

The next step is to note the number of factors common to each pair of variates, as in this table

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	—	2	5	0	5	2	5	1	5	4
x_2	2	—	5	1	3	5	5	0	2	4
x_3	5	5	—	1	7	8	12	1	8	10
x_4	0	1	1	—	1	1	1	0	0	1
x_5	5	3	7	1	—	4	7	1	5	6
x_6	2	5	8	1	4	—	9	0	5	8
x_7	5	5	12	1	7	9	—	1	9	11
x_8	1	0	1	0	1	0	1	—	1	1
x_9	5	2	8	0	5	5	9	1	—	8
x_{10}	4	4	10	1	6	8	11	1	8	—

From these, and from the total number of factors both specific and group in each variate, can be found the correlation which would occur between the variates were we to throw dice, one to each factor, and repeat the throwings a large number of times. The formula is

$$r = \frac{\text{Number of common factors}}{\text{Geometrical mean of totals}}.$$

This formula is applicable not only to variates formed by the addition of dice, but to variates which are any function of the factors or elements, provided that the form of the function is the same in each variate, and that the standard deviation is the same for each element or factor*. We thus obtain the following table of theoretical correlation coefficients:

* G. H. Thomson, *loc. cit.* p. 275. It can readily be deduced from Bravais, *Mémoires de l'Institut de France*, 1846, IX. eqn. 28.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Totals
x_1	—	0.22	0.32	0.00	0.44	0.17	0.31	0.18	0.46	0.32	2.42
x_2	0.22	—	0.39	0.18	0.29	0.47	0.35	0.00	0.20	0.35	2.45
x_3	0.32	0.39	—	0.10	0.39	0.44	0.48	0.12	0.47	0.51	3.22
x_4	0.00	0.18	0.10	—	0.14	0.13	0.10	0.00	0.00	0.12	0.77
x_5	0.44	0.29	0.39	0.14	—	0.30	0.38	0.16	0.40	0.41	2.91
x_6	0.17	0.47	0.44	0.13	0.30	—	0.47	0.00	0.38	0.53	2.89
x_7	0.31	0.35	0.48	0.10	0.38	0.47	—	0.11	0.51	0.54	3.25
x_8	0.18	0.00	0.12	0.00	0.16	0.00	0.11	—	0.17	0.15	0.89
x_9	0.46	0.20	0.47	0.00	0.40	0.38	0.51	0.17	—	0.58	3.17
x_{10}	0.32	0.35	0.51	0.12	0.41	0.53	0.54	0.15	0.58	—	3.51

If we wished to obtain experimental values of these, dice would have to be thrown, one to each factor. The die corresponding to the group factor called "Ace" would have its score counted into every variate containing the group factor in question. The dice representing the specific factors would, of course, only be counted into the one variate in which they occur.

The last column of the preceding table gives the total correlation of each variate with all the others, found by adding the rows of the square table. Rearrange now the sequence of the variates in the order of magnitude of these totals*, and we obtain the following table:

	x_{10}	x_3	x_9	x_7	x_5	x_6	x_2	x_1	x_8	x_4
x_{10}	—	0.51	0.58	0.54	0.41	0.53	0.35	0.32	0.15	0.12
x_3	0.51	—	0.47	0.48	0.39	0.44	0.39	0.32	0.12	0.10
x_9	0.58	0.47	—	0.51	0.40	0.38	0.20	0.46	0.17	0.00
x_7	0.54	0.48	0.51	—	0.38	0.47	0.35	0.31	0.11	0.10
x_5	0.41	0.39	0.40	0.38	—	0.30	0.29	0.44	0.16	0.14
x_6	0.53	0.44	0.38	0.47	0.30	—	0.47	0.17	0.00	0.13
x_2	0.35	0.39	0.20	0.35	0.29	0.47	—	0.22	0.00	0.18
x_1	0.32	0.32	0.46	0.31	0.44	0.17	0.22	—	0.18	0.00
x_8	0.15	0.12	0.17	0.11	0.16	0.00	0.00	0.18	—	0.00
x_4	0.12	0.10	0.00	0.10	0.14	0.13	0.18	0.00	0.00	—

Here the tendency to hierarchical order is quite noticeable. This particular example is purposely chosen from among a number calculated, as being that which shows the *least* hierarchical tendency. Even here however there is clearly a general lowering of the coefficients as we pass either along a row or down a column. The columnar correlation is high for the first few columns. For the columns headed x_{10} and x_3 it is 0·97, and as far down as the columns headed x_5 and x_6 it is still 0·65. If these theoretical numbers were blurred by experimental error, they might well be claimed as having come from a perfect hierarchy by the criteria in vogue. Other hierarchies chosen at random from those formed in the above manner, show still more perfect hierarchical order, and some-

* A convenient plan, but of no theoretical significance.

times the hierarchical order is almost quite perfect, as in the example given in the *British Journal of Psychology*, 1919, IX. p 343

(3) APPLICATION OF THE "CRITERION" TO THESE CASES, APPARENTLY PROVING THE PRESENCE OF A GENERAL FACTOR

The values of the correlation coefficients given in the above table are of course the *real* values To obtain experimental values of these, dice were thrown, one die to each Group or Specific Factor, and the whole repeated 30 times, analogous to experiments on 30 subjects*

From the dice scores the observed correlations between the variates can be calculated Using the product-moment formula we obtain the set of values in this table

The Observed Hierarchy

	x_{10}	x_5	x_9	x_6	x_7	x_1	x_3	x_2	x_4	x_8
x_{10}	—	66	61	71	69	45	52	37	24	− 07
x_5	66	—	67	57	52	45	36	33	25	19
x_9	61	67	—	49	58	58	·40	28	·10	03
x_6	71	·57	49	—	57	28	42	58	− 01	01
x_7	·69	52	58	57	—	·33	58	43	− 11	02
x_1	45	45	58	28	·33	—	59	23	27	− 06
x_3	52	36	40	42	58	59	—	23	− 14	05
x_2	·37	33	28	58	·43	23	23	—	04	− 14
x_4	·24	25	10	− 01	− 11	·27	− ·14	04	—	− 10
x_8	− 07	19	03	01	02	− 06	05	− 14	− 10	—

The pairs of columns which pass the Hart and Spearman correctional standard give the following values:

Columns passing standard	Observed columnar correlation R	True columnar correlation	The Hart and Spearman corrected columnar correlation R'
3 and 6	0·72	0·88	0·87
3 „ 7	0·83	0·99	0·98
3 „ 9	0·89	0·87	1·14
3 „ 10	0·73	0·98	0·85
6 „ 7	0·94	0·88	1·08
6 „ 9	0·78	0·62	0·92
6 „ 10	0·83	0·83	0·90
7 „ 9	0·81	0·86	0·93
7 „ 10	0·84	0·99	0·91
9 „ 10	0·89	0·84	1·04
Means	**0·83**	**0·87**	**0·96**

True mean columnar correlation of the whole table and not merely of the pairs of columns selected by the correctional standard } **0·59**

* G. H Thomson, *Biometrika*, 1919, XII. pp. 355—366, where the full details of the dice throws are given, corrected and continued in 1923, XV. pp. 150—160. Another example is also given there Sections 3 and 4 of the present chapter consist largely of extracts from *Biometrika*, where the diagrams 28 and 29 appeared.

Dr Hart and Professor Spearman would therefore claim the hierarchy as being a sample of a perfect one The true mean columnar correlation for the whole table is 0·59, the Hart and Spearman correctional standard selects pairs of columns whose true mean columnar correlation is 0·87, and the mean value of these when corrected according to their formula rises to 0 96 This example goes far towards shaking confidence in their criterion.

(4) THE ERRONEOUS NATURE OF THE HART AND SPEARMAN CRITERION*

The inaccurate and exaggerated estimates of hierarchical order which are given by this "criterion" arise chiefly from two causes, (1) the erroneous assumption that ρ' and ϵ are uncorrelated (see p 170), and (2) the action of the "correctional standard" We shall consider these in turn.

Consider the formula for the standard deviation of a correlation coefficient, viz.

$$\sigma_r = \frac{1 - r^2}{\sqrt{N}},$$

where N is the number in the sample. It follows from this that the larger correlation coefficients will probably have the smaller sampling errors e, disregarding the sign of e for the moment

But these signs of the quantities e are not likely to be indiscriminately positive and negative. On the contrary, they will have a tendency to be either all positive or all negative, if, as is the case in most of the columns of coefficients considered by Professor Spearman, the correlations in the square table are mainly positive. The errors in the correlation of a variate x_1 with a variate a are themselves correlated with the errors in the correlation of the variate a with another variate x_2, according to the formula†

$$r_{x_1x_2} = \frac{r_{x_1a} r_{x_2a} (1 + 2r_{x_1x_2} r_{x_2a} r_{x_1a} - r_{x_1x_2}^2 - r_{x_2a}^2 - r_{x_1a}^2)}{2(1 - r_{x_1a}^2)(1 - r_{x_2a}^2)}$$

That is, the correlation of the sampling errors of r_{x_1a} with the sampling errors of r_{x_2a} depends chiefly upon $r_{x_1x_2}$. To illustrate, let us take three

* See note on p 192.

† Karl Pearson and L N G Filon, "On the Probable Errors of Frequency Constants," *Phil. Trans of the Royal Soc.* 1898, CXCI A, eqn 37.

correlations from an experiment in psychology, carried out by Mr Wyatt*. If we let

x_1 be the mental test "Rearranged Letters,"
x_2 ,, ,, ,, "Missing Digits,"
a ,, ,, ,, "Analogies,"

the values there found were

$$r_{x_1 a} = 0{\cdot}63,$$
$$r_{x_2 a} = 0{\cdot}61.$$

Then by the above formula the correlation of the errors of these two coefficients depends chiefly upon $r_{x_1 x_2}$, whose measured value is 0·63. Using the full formula, and employing the measured values in default of the true ones, the correlation between $r_{x_1 a}$ and $r_{x_2 a}$ turns out to be ·47. It is therefore (to an extent indicated by this value) probable that they are either both too large or both too small The same argument holds, in varying degrees, for the other correlations all over Mr Wyatt's table, which are all positive They all have a tendency to be either all too large or all too small. in other words, the e's tend to be all of the same sign. The relationship between the correlation coefficients of a column, and their errors, can therefore be summed up in the following table, in which the symbol $|e|$ denotes the magnitude of e regardless of sign.

r'	$\|e\|$	ρ'	ϵ or ϵ		$\rho'\epsilon$ or $\rho'\epsilon$	
large	small	+	−	+	−	+
		+	−	+	−	+
		+	−	+	−	+
		−	−	+	+	−
		−	+	−	−	+
		−	+	−	−	+
small	large	−	+	−	−	+
			$S(\rho'\epsilon) =$		− or +	

The first column shows the true correlations r' arranged in order of magnitude. The second column expresses the fact that the sampling errors on any occasion will probably be arranged in the reverse order of magnitude, disregarding their signs. The third column shows the correlation coefficients measured from their mean. The upper ρ''s are then positive, and the lower negative, and also, what is not shown in the

* Stanley Wyatt, "The Quantitative Investigation of Higher Mental Processes," *Brit. Journ. Psychol* 1913, VI. p 131

table, the absolute values increase upwards and downwards from the point where the signs change The fourth (double) column shows the probable arrangement of the signs of the quantities ϵ If the e's are all tending to be positive, then the left-hand member of the double column gives the arrangement, while if the e's all tend to be negative, the other member of the double column does so As shown in the last (double) column, therefore, the quantities $\rho'\epsilon$ tend either to be nearly all negative or nearly all positive. For a very small sample the signs of $\rho'\epsilon$ will no doubt be quite irregularly arranged. But with such a small sample, even if ρ' and ϵ were really uncorrelated, it would be most unlikely for $S(\rho'\epsilon)$ to be negligible As the sample increases the signs tend to settle down to the above arrangement, and $S(\rho'\epsilon)$ does not tend to disappear compared with $S(\epsilon\epsilon)$, but only to take on one or other of alternative values. It will only be zero when *all* the errors are zero, i.e. when *no* corrections are needed to R'. The distribution of $S(\rho'\epsilon)$ about zero in a number of samples of the same size will not, that is, show a maximum at zero, but a minimum, as is shown qualitatively in Fig. 28.

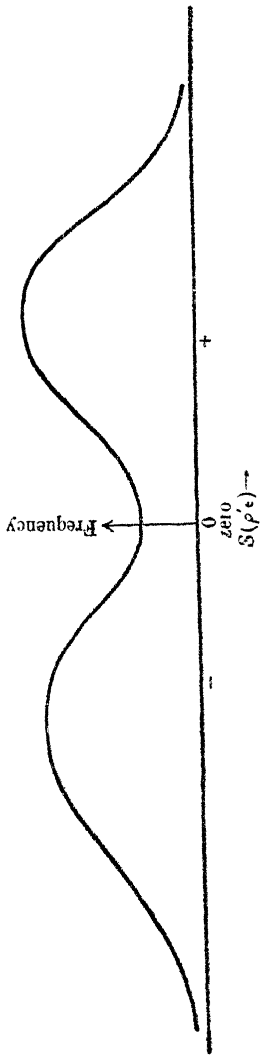


Fig. 28

If, in fact, the actual value of $S(\rho'\epsilon)$ is calculated in cases where the *true* correlations are known, it is frequently found to be greater than the quantities $S(\epsilon\epsilon)$ which are left in the expression.

The other approximations made in obtaining the criterion do not appear to be so erroneous as this one, though their cumulative effect may explain some anomalies. Leaving them on one side let us consider the "correctional standard" required by Dr Hart and Professor Spearman before they admit any pair of columns. It is this correctional standard, combined with the peculiar distribution of R', which chiefly is responsible for the exaggeration of perfection produced

by this criterion, and for the regularity with which an average value of unity is arrived at.

Let us examine first the actual distribution of the Hart and Spearman R' in a psychological hierarchy, viz. that of Wyatt already referred to, and calculate R' not only for those columns which pass the correctional standard, but also for other pairs of columns What we find is that its value rises as we descend the hierarchy, rushing asymptotically to infinity, remaining for a time imaginary, and then returning Specimen values from Mr Wyatt's hierarchy are given.

Pairs of columns	*Values of the Hart and Spearman Criterion*	
Analogies and Wordbuilding	0·93	
Completion and Wordbuilding	0·97	
Completion and Part-wholes	1·05	Passed by the correctional standard
Wordbuilding and Part-wholes	0·99	
Part-wholes and Memory (delayed)	0·92	
Rearranged letters and Missing digits	1·17	
Wordbuilding and E R Test	1·26	
Sentence construction and Fables	1·33	
Rearranged letters and E R Test	Practically infinity	
Nonsense syllables and dissected pictures	Imaginary	
Crossline test and Letter Squares	0·35, *a meaningless value*, both factors in the denominator being now negative	

Expressed in diagrammatic form this and similar calculations lead to the conclusion that in actual practice the criterion is distributed as in Fig. 29, where the curve is to be understood as a "best fitting" curve among the values of R' scattered, with a very considerable dispersion, on both sides of it. The line, in fact, ought to be a broad smudge.

Now clearly, with a distribution of this sort, it is very important that the boundary between the values that are to be rejected and those that are to be accepted should be chosen with the greatest care, and not arbitrarily but scientifically. Either sound theoretical reasons should be given for the choice of the correctional standard, or the choice should be based empirically on experiments in material where the truth is known *a priori*, as in the above dice experiments. For obviously, by moving this boundary, we can make the final average take on almost any value. Another point is that the criterion rushes to infinity at such speed that its probable error must be enormous. Dr Hart and Professor Spearman, however, give no reasons for their choice of this particular standard, upon which depends so much the values they obtain. The standard which they thus arbitrarily adopt begins admitting the criteria at just such a distance above unity as to balance the cases which give a criterion below unity, and entirely explains the remarkable unanimity with which this average value unity is obtained by them in their calculations.

In other words, the remarkable regularity with which this criterion gives the value unity is not a property of the investigated correlation coefficients at all, but is a property possessed by the criterion itself, due to errors and the action of the "correctional standard."

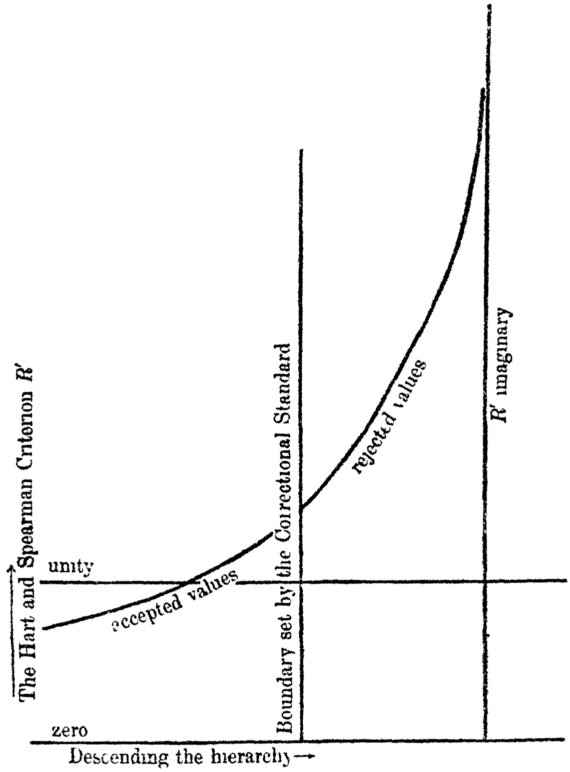


Fig 29

In the writers' opinion the work outlined in this chapter finally proves the invalidity of Professor Spearman's mathematical argument in favour of the Theory of Two Factors If this be so that theory returns to the status of a possible, but unproven, theory.

(5) HIERARCHICAL ORDER THE NATURAL ORDER AMONG CORRELATION COEFFICIENTS

The fact is that hierarchical order, which Professor Spearman was the first to notice among correlation coefficients, is the natural relationship among these coefficients, on any theory whatever of the cause of the correlations, excepting only theories specially designed to prevent

its occurrence. It is the *absence* of hierarchical order which would be a remarkable phenomenon requiring special explanation, its presence requires none beyond what is termed chance.

An analogy from the simple repeated measurements of a linear magnitude may help to illustrate this Indeed it is rather more than an analogy, being in fact the same phenomenon in its simplest terms and dimensions It is well known that many measurements of the same quantity, made with all scientific precautions, under apparently the same conditions, and with an avoidance of all known sources of error, nevertheless do not give a number of identical values. The values are all different, but are not without law and order in their arrangement They are grouped about a centre from which the density decreases in both directions, and it is found that this grouping is for most practical purposes closely represented by the Normal or Gaussian Curve of Error (or one of the more general Pearsonian Curves)

Experimenters are not surprised to find their data obeying the Probability Law, nor do they require a special theory to explain it. On the contrary, it is the departures from this Law which if wide would require special investigation, and if confirmed would require a special theory. In the same way hierarchical order among correlation coefficients should not cause surprise, though any marked variation from this order would demand investigation.

Measured correlation coefficients are themselves correlated, and n coefficients form an n-fold or n-dimensional correlation-surface. The particular and convenient form of tabulation of correlation coefficients adopted by Professor Spearman and followed by most other psychological workers brings to light, in the form of "hierarchical order," one of the properties of this correlation-surface of the correlations.

In an article entitled "On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation," in the *Phil. Trans.* 1898, CXCI. A, pp 229—311, Professor Pearson and Mr Filon give the following formulae.

$$R_{r_{12}r_{13}} = r_{23} - \frac{r_{12}r_{13}\,(1 + 2r_{12}r_{23}r_{31} - r_{12}^2 - r_{23}^2 - r_{31}^2)}{2\,(1 - r_{12}^2)\,(1 - r_{13}^2)},$$

$$R_{r_{12}r_{34}} = \frac{\left\{\begin{matrix}(r_{13} - r_{12}r_{23})\,(r_{24} - r_{23}r_{34}) + (r_{14} - r_{13}r_{34})\,(r_{23} - r_{21}r_{13}) \\ + (r_{13} - r_{14}r_{43})\,(r_{24} - r_{21}r_{14}) + (r_{14} - r_{12}r_{24})\,(r_{23} - r_{24}r_{43})\end{matrix}\right\}}{2\,(1 - r_{12}^2)\,(1 - r_{34}^2)},$$

so that, as they say, "errors in the correlations of a first organ with a second and a third have a correlation themselves of the first order,"

and "errors in the correlation of two organs and in the correlation of a second two have only correlation of the second order"

Suppose now that the correlations among a number of variates taken in pairs are really all the same, and positive, and in a sample let the observed value of r_{38} be the highest observed value

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9 . . .
x_1			h					h	
x_2			h					h	
x_3	h	h	—	h	h	H_2	h	H_1	h . . .
x_4			h					h	
x_5			h					h	
x_6			H_2			—		H_3	
x_7			h					h	
x_8	h	h	H_1	h	h	H_3	h	—	h . . .
x_9			h					h	
x_{10}			.					.	
.			.					.	

H_1 = highest correlation H_2 = second highest, and so on h = tendency to be high.

Then, because of the above theorems of Pearson and Filon, the rows and columns x_3 and x_8 will probably contain more total correlation than do the others, and the second highest correlation will probably be in one of these Let it be r_{36}. Then, after the rows x_3 and x_8 the row x_6 will probably contain many high correlations and r_{86} will probably be the third highest coefficient, because it is a *node* where two *ridges* of high correlation cross. If it is, then the hierarchy so far is excellent, as can be seen on rearranging the square table so as to bring x_3, x_8, and x_6 to the head*.

Where now is the fourth highest r likely to be found? Somewhere among those marked h for high, and slightly more probably among the r's of the row x_3. If to avoid further rearranging we take it to be r_{31}, then r_{81} is most probably the fifth, and r_{61} the sixth, because they are nodes. And so on. In fact it is clear that when the r's are really equal

* See next page.

to one another, then sampling the population will give a set of observed r's which are arranged not in haphazard, but in hierarchical order.

	x_3	x_8	x_6	x_1	x_2	x_4	x_5	x_7	.	.	.
x_3	—	H_1	H_2	h_4	h	h	h	h	.	.	.
x_8	H_1	—	H_3	h_5	h	h	h	h	.	.	.
x_6	H_2	H_3	—	h_6	h	h	h	h	.	.	.
x_1	h_4	h_5	h_6								
x_2	h	h	h								
x_4	h	h	h								
x_5	h	h	h								
.	.	.	.								
.		.	.								

In the experiment described above with cards and dice, where hierarchical tendency was found to be produced by group and specific factors without any general factor, there was, however, no question of sampling the *population*. The hierarchical order already appears in the theoretical coefficients. There is here, however, another kind of sampling present, viz. sampling of the *elements* which make up the variates.

Let us suppose, instead of deciding the numbers of groups and specific elements as we did by drawing cards, we had dipped into an infinite bag containing black balls and white balls, the former representing group factors and the latter specific factors. Then the most probable event would have been that the proportion of group factors and specific factors in each variate would have been the same as the proportion in which the balls occurred in the bag If, in addition, we assume the samples drawn to be all the same size, then the most probable result of the whole experiment would have been obtaining all the correlations equal.

That they do not come out equal may be regarded as due to the sampling. The samples vary in size, the proportion of group factors varies from sample to sample, and the distribution of the individual group factors among the several variates departs from the most probable distribution. From this point of view the departure of the correlation coefficients from equality is due to errors of sampling, and the application of the theorem of Pearson and Filon would lead us to expect what as we have seen does actually occur, namely, hierarchical order.

The experiment with the bag of balls would as a matter of fact not produce as great a departure from equality of correlation coefficients as is found in practice in experimental psychology, or as was found in our form of the experiment with card drawing This in the latter case is because the cards give a greater variation of the proportion of group to specific factors than would be found with the bag of balls. And this is also the case in mental tests, which are not *chance* samples of the mental elements, but are carefully chosen so as to measure different kinds of activity.

This application of Pearson and Filon's theorem (which contemplates only sampling errors in the ordinary sense) to the changes in correlation produced by sampling the underlying elements, is no doubt somewhat novel, and may appear to be a difficulty.

Further consideration leads to the following resolution of the difficulty*.

Suppose that n variates (in our work the scores in mental tests) are so connected by factors that the correlations are all equal and positive Then let a small sample of the population be taken The *observed* correlations will show departures from equality, and will be found to be in hierarchical order. This hierarchical order is due to sampling the population

Now consider why the correlations do not come out at their true values. They give of course the true values *for the sample.* The reason of their departing from the true values of the whole population is that (*a*) some of the factors which really are links between the variates (the mental activities) happen to have remained steadier than usual during the sample. In the limit a factor might happen to retain exactly the same value through the various individuals of the sample. That is, some of the linking factors do not in reality come into action, or not in their full force, (*b*) on the other hand, some factors which are really different and unconnected may happen by chance to rise and fall together, through the sample, and more or less to act as one. That is, fictitious linking factors are created, which would disappear with a larger sample.

Clearly therefore a hierarchy of correlation coefficients, caused by sampling the population, is due to chance having caused a change in the apparent factors acting. It follows that if we make a real change in the factors acting, we shall get a hierarchy, and this is what we do when we choose the mental tests to be employed in any research. Each mental test is a test of a sample of abilities.

* The following is quoted from p. 182 *et seq* of *Psychol. Review*, 1920, xxvii No 3.

The laws governing the correlation of correlation coefficients which vary because of sampling the population can, in fact, be applied without hesitation to the relationships between "true" correlations in the whole of any population simply because any such population is itself a sample. English grammar school boys of 12 are themselves a sample of a larger boyhood, the whole human race indeed is a sample of "what might have been," selected by the struggle for survival.

The whole question clearly has philosophical bearings on the degree of reality of causal connections; for on this view those chance links in a small sample which were a few paragraphs ago termed "fictitious links, which would disappear with a larger sample," do not differ except in degree from the "real" causal links which we only term real because they persist throughout the largest sample with which we are acquainted.

In another direction there are connections with the difference, which is one of degree only, between what is called "partial" correlation and "entire" correlation*.

The conclusion to be drawn is that hierarchical order is the natural order to expect among correlation coefficients, on a theory of chance sampling alone, and that therefore, by the principle of Occam's razor, its presence† cannot be made the criterion of the existence of any special form of causal connection, such as is assumed in the Theory of Two Factors.

(6) THOMSON'S SAMPLING THEORY OF ABILITY‡

In place therefore of the two factors of that theory, one General and the other Specific, Thomson prefers to think of a number of factors at play in the carrying out of any activity such as a mental test, these factors being a sample of all those which the individual has at his command

The first reason for preferring this theory is that of Occam's razor. It makes fewer assumptions than does the more special form of theory It does not deny General Ability, for if the samples are large there will of course be factors common to all activities On the other hand it does not assert General Ability, for the samples may not be so large as this, and no single factor may occur in every activity. If moreover a number of factors do run through the whole gamut of activities, forming a General Factor, this group need not be the same in every individual.

* See Karl Pearson, "On the Influence of Natural Selection on the Variability and Correlation of Organs," *Phil Trans Roy Soc. London*, 1902, cc. A, pp 1—66, Godfrey H. Thomson, "The Proof or Disproof of the Existence of General Ability," *Brit Journ Psychol.* 1919, IX. pp. 321—336

† See ch XI, and meanwhile read "its presence, unless in a degree of perfection greater than chance would explain" Note added 1924.

‡ *Psychol. Review*, 1920, XXVII. p 183

In other words General Ability, if possessed by any individual, need not be psychologically of the same nature as any General Ability possessed by another individual Everyone has probably known men who were good all round, but Jones may be a good all round man for different reasons from those which make Smith good all round.

The Sampling Theory, then, neither denies nor asserts General Ability, though it says it is unproven. Nor does it deny Specific Factors. On the other hand it does deny the absence of Group Factors It is this absence of Group Factors which is in truth the crux of Professor Spearman's theory, which is not so much a theory of general ability, or a theory of two factors, as a Theory of the Absence of Group Factors And inasmuch as its own disciples have begun to require Group Factors to explain their data, its distinguishing mark would appear in any case to be disappearing

Such Group Factors as are admitted by Professor Spearman are of very narrow range, and are mutually exclusive, that is they do not overlap. Both these points follow from the sentence used in the 1912 article with Dr Hart, where it is said that, in the case of performances too alike, "when this likeness is diminished, or when the resembling performances are pooled together, a point is soon reached where the correlations are still of considerable magnitude, but now indicate no common factor except the General one."

Since this point is soon reached, the Group Factors must be narrow in range. Since pooling a few performances will obliterate any Group Factors, they must be exclusive of one another. For if A, B, C and D are four tests, in which A and B have a Group Factor common to them, and C and D another, then of course by pooling A with B and also C with D we can obtain two pools AB and CD which have no link. But if A, B and C have one Group Factor, and C and D have another, then these Group Factors cannot be separated into Specific Factors. In fact, a Specific Factor is a separated Group Factor, and Professor Spearman's theory asserts that Group Factors, if any, are separable and mutually exclusive This is a great stumbling-block in the way of the acceptance of the Theory of Two Factors, unless perhaps "Specific Factor" is interpreted in the way suggested later.

It is a fact which will be admitted by most that the same activity is not performed in the same way by different individuals, even though they are equally expert Not only are Specific Factors therefore required by this theory for every separate activity, excluding only any which are very closely similar, but also Specific Factors of different psycho-

logical natures are required for each individual. Further, the same individual does not always perform the same activity in the same way. A man using an ergograph will, as he tires, begin to employ muscles other than those naturally used at the outset. When we are returning from a cycle ride muscles are used in a different manner from the style adopted at the start, indeed sometimes deliberate changes are made to give relief. And in the same way a mental task is performed by different methods at different times. Does this then mean a different Specific Factor for each way of doing a task? All these difficulties appear to argue against the Theory of Two Factors, and seem to be considerably cleared up by the Sampling Theory.

Finally, the Sampling Theory appears to be in accordance with a line of thought which has already proved fruitful in other sciences. Any individual is, on the Mendelian theory, a sample of unit qualities derived from his parents, and of these a further sample is apparent and explicit in the individual, the balance being dormant but capable of contributing to the sample which is to form his child. It seems a natural step further to look upon any activity carried out by this individual as involving a further sample of these qualities.

(7) THE DIFFICULTY OF "TRANSFER OF TRAINING"

Although Professor Spearman's Theory of Two Factors has been chiefly based by him on the line of argument which, it is suggested, has now been proved invalid, viz. the "hierarchy" argument, yet there is another and powerful form of reasoning which can be brought to its support, based upon the fact that, according to some experimenters, improvement in any activity due to training does not transfer in any appreciable amount to any other activity, except to those very similar indeed to the trained activity. And even those workers who do not agree that this is an experimental fact are usually content to take a defensive attitude and say that transfer is not disproved. Few if any will say that it is proved.

This certainly seems to point to the absence of Group Factors, and to support Professor Spearman's theory, which only needs to add to itself the assumption that the Specific Factors are, while the General Factor is not, capable of being improved by training, to fit the case admirably. Of course, if transfer really occurs, the argument proves the opposite. And although psychological experiment points on the whole to the absence or the narrowness of transfer, yet popular opinion among business men, schoolmasters, and others, is in favour of transfer

to a considerable extent. Assuming no transfer, however, how can the Sampling Theory, with its numerous Group Factors, explain this?

It is necessary to assume that the Group Factors are all unimprovable or only slightly improvable by training, though they may change with the growth and development of the individual. The improvement which certainly takes place when we practise any activity is due, it may then be assumed, not to improvement in the elemental abilities which form the sample, *but to a weeding out*, and selection of these The sample alters, mainly no doubt is diminished, though additions are also conceivable. It becomes a more economical sample, and waste of effort in using elements which are unnecessary is avoided. Improvement in any mental activity may on this view be compared with improvement in a manual dexterity, in which it is notorious that the improvement consists largely in the avoidance of unnecessary movements

When another activity is then attempted, the elemental factors are just the same as they would have been had the practice in the first activity not taken place. The new activity will be performed by a new group of factors, which sample will as in the first case be in the beginning wasteful and will include many unnecessary elements. Transfer of improvement gained in the first activity will therefore not take place except in so far as the second activity is recognised as a mere variant of the original one, in which case the weeding out process which has taken place in the first case may be done at the very first attempt, at any rate to some extent.

To use another analogy, the improvement which takes place when a football team practises playing together for a series of matches is due more to team work than to individual improvement. A new team, even though it contain a large proportion of players from the first team, will not have this unity of action There will be little transfer of improvement.

According to the view here developed, it is the weeding out of the sample of elemental abilities which is specific The team work is specific, though the players play for several clubs. This would appear to enable a reconciliation to be effected between the almost universal belief in "types" of ability (to which Professor Spearman refers) and the experimental facts concerning both correlation and transfer. If there be a General Factor at all, it might be the power to shake down rapidly into good team work, in a word, educability. But there seems no objection to assuming that this, instead of being a General Factor, is a property of each elemental factor, varying from factor to factor

To sum up this section if transfer of training really does not occur to any great extent, then it has to be admitted that the Theory of Two Factors readily explains this But the Sampling Theory can also do so, in a manner which is perhaps not so easy to set forth, but which nevertheless appears to be more illuminating and less artificial than the alternative theory

(8) CONCLUSIONS

Professor Spearman's Theory of Two Factors, which assumes that ability in any performance is due to (*a*) a General Factor and (*b*) a Specific Factor (Group Factors being absent, or at any rate very narrow in range and mutually exclusive), is based chiefly on the observed fact that correlation coefficients in psychological tests tend to fall into "hierarchical order" It has been shown, however, that the criterion adopted for evaluating the degree of perfection of hierarchical order present is untrustworthy and has led to over-estimation. Such hierarchical order as is actually present is in fact the natural thing to expect, and it is the absence of such which should occasion surprise. The proof of the Theory of Two Factors which is based on the presence of hierarchical order therefore falls to the ground. The theory remains a possible explanation of the facts but ceases to be the unique explanation. As an alternative theory Thomson has advanced a Sampling Theory of Ability, in which any performance is considered as being carried out by a sample of Group Factors This theory is preferred because it makes fewer and less special assumptions, because it is more elastic and wider, and because it is in closer accord with theories in use in biology and in the study of heredity.

Note, 1924 Mr H G Stead has made the views expressed in pp 179—183 the subject of experiment (*Journ Roy Stat Soc* 1923, LXXXVI p 412) and obtains results which do not agree with the suggestion that the e's tend to be of the same sign (p 180) or with the suggested distribution of $S\ (\rho'e)$ (diagram p 181) As to the first point, it would seem that the tendency must be there if Pearson and Filon's formula is correct (p 179) *provided the correlations are positive* But the tendency will be the less, the smaller the correlations; and we understand that many of Mr Stead's correlations were very small, which possibly explains the discrepancy As to the distribution of $S\ (\rho'e)$ we think (though he does not agree) that Mr Stead has plotted the wrong quantity The diagram on p 181 does not mean that a number of values of $S\ (\rho'e)$ from different columns of one experiment will have two maxima, but that one value of $S\ (\rho'e)$, when obtained many times in many experiments, will show that property.

With the second part of the argument, on pp 182 and 183, Mr Stead's results and his expressed opinions are in complete agreement.

CHAPTER XI

THE PRESENT POSITION (1924)

It appears desirable to utilise the remaining available pages for a statement of the present position of the controversy which forms the subject of the two preceding chapters, especially as in discussion points of difference are apt to be magnified and points of agreement lost sight of.

The term "general intelligence" or "general ability" is liable to have two distinct meanings On the one hand it may be a statement of a fact, on the other an explanation of that fact The fact which makes the term a necessary and a useful one is that a man who is good at one kind of mental work is usually above the average in others*. In technical language, most measures of correlation between various mental tests, or between various school and university subjects, are positive, and many are high. Though some are low, few are negative When this is denied, it is generally on the strength of a number of individual cases where marked ability is found in one subject but not in another. These are, however, swamped by the much larger number of cases in agreement with the principle Because of this fact of predominant positive correlation, it is possible, after administering an intelligence test lasting one or two hours, to predict an individual's performance in various mental activities with more or less probability, though never, of course, with absolute certainty. If the known correlation between the test and a certain other activity is r, then an individual who deviates d from the average in the test (in sigma units) will deviate rd from the average in that activity *most probably*. In practice however such individuals who deviate d in the test will not all be exactly at rd in the other activity, but will be scattered about it. And that scatter will be less than the scatter of an unselected group in the proportion $k:1$, where $k=\sqrt{(1-r^2)}$. The test by its constituent elements probes the mind at a number of different points and strikes an average, just as one finds the depth of a lake by plumbing it at various points It is then possible to make a prediction of its depth at some other point The average of the recorded depths would be one such prediction, and this is analogous to using the lumped score, or the I Q., in say a Binet test. A contoured map of the

* See G. H Thomson, "The Nature of General Intelligence and Ability," *Brit. Journ. Psychol.* 1924, XIV. p 229 and other articles of the same symposium at the VII International Congress of Psychology, Oxford, 1923, by Claparède and Thurstone, in the same *Journal.*

lake bottom, made from the recorded depths, would enable a better prediction to be made, analogous to using "profiles" in testing, though in testing we do not know the relationship of our elements as we know the spatial relationship of the points of a lake

There is no doubt that such predictions become increasingly precarious as the general similarity between the performances decreases From an intelligence test a prediction of some value could be made of a man's ability to learn to understand the theory of relativity: but not a prediction of his ability to throw a cricket ball. And a prediction of his probable ability, after training, in playing the piano would be between the two. Probably it was because the performances considered were far apart that Thorndike in his earlier work was led to say that one could almost believe that there was nothing whatever common to them Since those days tests have been more and more confined to abstract activities and correlations have risen. and they have also risen because methods of measurement have become more exact.

With this general *fact* of positive correlation the controversy under consideration has nothing to do. It is concerned with two somewhat different *ways of explaining* that fact

In the factual meaning, the term general intelligence is non-controversial. But it has come also to have another and more technical meaning, as the name of a general factor g which is supposed, on Professor Spearman's theory, to play a part in all our activities, to be (when they are sufficiently dissimilar) the sole cause of correlation between them and together with another factor specific to each activity to be the complete determiner of performance in that activity. Thomson's theory would explain correlations by assuming that each activity is a sample of many factors, much smaller than the factors contemplated by Spearman. It is atomic in its tendency, as the Mendelian theory is atomic.

As regards this wider area of discussion, it must be remembered that Thomson has never questioned the possibility of the Theory of Two Factors or claimed that he had disproved it. In his first paper* in 1916 he wrote: "The object of this paper is to show that the cases brought forward by Professor Spearman in favour of the existence of General Ability (g) are by no means crucial. They are, it is true, not inconsistent with the existence of such a common element but neither are they inconsistent with its non-existence." In a paper published in 1919† he wrote: "The result of the investigation is to confirm the statement already made that there are many theories *in addition to* that of Professor

* *Brit. Journ. Psychol.* VIII. p. 271. † *Ibid* IX. p. 273.

Spearman, which will explain such hierarchical order as is actually found ... The essence of all these theories is stated as conclusion" That conclusion was an early statement of his Sampling Theory from which the distinction there made of elements at two levels has since been dropped as unnecessary. And in 1920 Thomson wrote* of his Sampling Theory, "It does not deny general ability, for if the samples are large there will of course be factors common to all activities"

Indeed the only acute point of dispute was about Spearman's method of supporting his explanation and that may now perhaps be set aside since Spearman is engaged in making available the correct criterion in place of the "substitute that could supply good approximations under certain circumstances but was liable to mislead under others†" A recent paper by Spearman and Holzinger is of considerable importance. In it these authors find the probable error of

$$F \equiv r_{13}r_{24} - r_{23}r_{14},$$

a quantity which must be zero if the Theory of Two Factors is to be confirmed They find

$$\sigma_F{}^2 = \frac{1}{N}\left[r_{13}{}^2 + r_{14}{}^2 + r_{23}{}^2 + r_{24}{}^2 - 2\left(r_{12}r_{13}r_{23} + r_{12}r_{14}r_{24} + r_{13}r_{14}r_{34} + r_{23}r_{24}r_{34}\right) + 4r_{12}r_{13}r_{24}r_{34}\right] + \frac{k_{13}{}^4k_{24}{}^4 + k_{23}{}^4k_{14}{}^4}{N^2}, \quad \text{where } k^2 = 1 - r^2.$$

An approximation got by replacing each r by their mean is

$$\sigma_F{}^2 = 4r^2(1-r)^2/N + 2k^8/N^2,$$

and in this, the term divided by N^2 can be neglected unless both N and the r's are small, and therefore

$$\sigma_F = 2r(1-r)/\sqrt{N}.$$

The Sampling Theory is not then a rival of the Theory of Two Factors The two theories may be true simultaneously. Both may be useful guides in threading a path through the difficulties of constructing and interpreting mental tests. Spearman's theory has been of incalculable value in this way to the English school, and no examples need be quoted The Sampling Theory is of use in other situations. By its imagery, for example, one can readily appreciate the fact, which in particular Clark L Hull has pointed out‡, that a new test to be added to a team ought to correlate *high* with the criterion and *low* with the present team: and one

* *Psychol Review*, 1920, XXVII. p 184.
† *Brit Journ Psychol* 1924, XV. p 17.
‡ *Journ. Educ. Psychol.* 1923, XIV. p 396.

can go on to see that a new test may *also* be useful if it correlates *low* with the criterion and *high* with the present team, provided one *subtracts* its weighted score*.

On the other hand, though both the Theory of Two Factors and the Sampling Theory may be true simultaneously they are by no means merely alternative ways of stating the same thing They only become identical if *perfect* hierarchical order among correlation coefficients is indeed a fact. In that case two factors are all that are needed to express completely any activity, one specific and one general, as Thomson has always admitted†, and as Spearman and following him Garnett have shown, and all that the Sampling Theory in that case does is to split up the "g" into an aggregate of smaller factors, a possibility which has always been borne in mind also by Spearman, who in his earliest enunciation (we believe) of his theory adds to the words "have in common one fundamental function," the further qualification "or group of functions‡."

If therefore it can be shown that the quantity $F \equiv r_{13}r_{24} - r_{23}r_{14}$ is, within the limits set by its probable error, significantly equal to zero, then for tests where this is so the Sampling Theory reduces to the Theory of Two Factors. It would not, of course, be sufficient to take cases where F had a large probable error· it would be necessary to include cases where every effort had been made to reduce the probable error, and show that F still did not depart significantly from zero. This task of surveying the available data by means of the quantity F we may perhaps assume Professor Spearman and his co-workers to be engaged in The final result would replace the Hart and Spearman survey of 1912 made by the former criterion R'.

As Spearman himself has shown§, even when $F = 0$ and the hierarchical order is perfect, it is *possible* to do without the General Factor and use only Group Factors. But there is in this case of course no doubt as to the order of preference of the two possibilities, the assumption of the presence of a General Factor being a perfectly natural one, and the assumption of the presence of a *peculiarly related* set of Group Factors being highly artificial

It is however otherwise when the hierarchical order is not perfect, when F departs significantly from zero. The Sampling Theory, which

* This latter point, which had apparently escaped general notice, was stated by Thomson in a paper read at the meeting of the New York branch of the American Psychological Association on Feb. 25th, 1924.

† See e g Garnett and Thomson, *Brit Journ. Psychol* 1919, IX. p. 367, in addition to earlier references.

‡ *Amer. Journ. Psychol.* 1904, XV. p. 284. § *Psychol. Review*, 1920, XXVII p 164.

includes all cases, then no longer reduces to the Theory of Two Factors, for it is then no longer possible to express each activity completely by a general and a specific factor A third category of factors necessarily enters, composed of Group Factors which are neither entirely specific nor entirely general, factors which run through some, though not through all of the tests The Theory of Two Factors must in this case be expanded into a Theory of Three Factors*, or better three *kinds* of factors, the three which Thomson has called General, Group and Specific Factors Each activity will then be determined by a number of factors, one the general factor, one the specific factor and the others group factors This is quite a different matter from the Theory of Two Factors, which by its very name denies Group Factors If hierarchical order departs significantly from perfection group factors are essential, and no rewriting of equations can eliminate them.

When hierarchical order was perfect we had in a certain sense the choice between a General Factor and Group Factors· but the Group Factors had to be related in a very artificial way and formed therefore a highly unplausible hypothesis When the hierarchical order departs significantly from perfection we *must* postulate *some* group factors, but we still have a choice whether we explain the remaining correlation by a General Factor or by Group Factors There is, however, this important difference from the case of perfect hierarchical order, that here the artificial and peculiar relationships between the Group Factors are no longer necessary, in proportion as the perfection of the hierarchy is relaxed for the departure of the Group Factors from those relationships is equivalent to supplying those Group Factors which must be present over and above a General Factor in order to explain the departure from $F = 0$. As long as correlations are all positive any one who wishes to do so may postulate a General Factor But unless hierarchical order is perfect he must also postulate Group Factors Now the Group Factors are the most general of the three categories of factors. The General Factor and the Specific Factors are each special cases of Group Factors. And therefore, although to postulate as large a General Factor as the correlations will allow is in one sense the simplest procedure by way of theory, in another and, we think, a wider sense, it is simpler to postulate only Group Factors, unless the approach to hierarchical order is so close that the Group Factors would be required to fulfil artificial and improbable conditions. How high an approach to hierarchical order is compatible

* This name was proposed by Dr Arthur S Otis, in a copy of a manuscript with that title sent to Thomson some time after the publication of Thomson's 1916 paper, but which Otis as far as we know has never published.

with unfettered Group Factors? This question Thomson first attacked by trying the special case of thirteen Group Factors (represented by cards of a suit) arranged in absolutely random order in ten imitation tests, and he invariably obtained* a very high degree of hierarchical order. In the same paper he gave the theoretical considerations (repeated on pp. 183–8) which led him to make the statement that "hierarchical order is the natural relationship among all correlation coefficients," meaning not *perfect* hierarchical order, but some degree of hierarchical order, and a larger degree than, he believed, Spearman had realised†. To this statement Thomson still adheres Udny Yule, in a critical notice of the 1921 print of this book‡, says that he parts company altogether with Thomson on this point To some extent this may be due to Yule's restriction of the term hierarchical order to cover only *perfect* hierarchical order. But in part, at least, it seems to be a real rejection of Thomson's idea that the laws of sampling apply to the *true* correlation coefficients and not merely to the errors For Yule says "It is not necessarily true that 'coefficients of correlation are themselves correlated' if by this is meant that r_{km} and r_{sm} are positively correlated for any pair of columns k and s. *Fluctuations of sampling* in the r's are so correlated" If Yule does reject this idea then he is, we think, mistaken. But it may well be that the fault is Thomson's lack of clearness in explaining, and a re-statement of what is contained in certain of Thomson's articles and repeated on pp. 187–8 of this book may perhaps be forgiven. The idea is that *true* correlation coefficients in a number of organs differ from level equality for the same kind of reason that sampled values of *truly equal* correlation coefficients differ from one another that when equal correlation coefficients appear to differ the reason is that the sampling has caused certain elemental factors to be present or absent simultaneously

* *Proc Roy Soc* 1919, xcv. A, p 400

† In the *Psychol Review*, 1914, xxi p 109, as a footnote to the phrase "the correlation between columns .must be zero," Spearman says. "A similar result ensues from the not unplausible hypothesis, that each performance depends on a randomly selected group of very numerous independent elements and that the correlation between any two performances is due to some of the elements happening to be common to both groups. For it could easily be shown that under these assumptions the correlation (compensated for sampling errors) between any two columns will tend to equal the correlation (uncorrected for attenuation) between the two performances from which the columns derive. .and both will average *little more than zero*" The proof of this was not given then, nor as far as we know has it been published since, though Spearman has referred to it again, e g in *Brit Journ. Psychol* 1916, VIII p. 283, and through Webb in "Character and Intelligence" (*Brit. Journ Mon. Supp.* 1915), pp. 57 and 82 Its publication seems to us very desirable, as Thomson's experience is that (supposing none of the elements are interference elements but all are positive) the hierarchical order will be much greater than this, at any rate if the elements are "all-or-none" in action

‡ *Brit. Journ. Psychol.* 1921, XII. p. 104.

as though identical, or has happened to select cases where identical elements usually present in both instances happen to be missing in one. that philosophically and mathematically the differences between really different correlation coefficients are also of this nature, for the variates are differing samples of the underlying elements. Thus if the correlations between three tests *a*, *b* and *c* were really equal, then in a given sample they would differ because (apart from errors of measurement) the bonds, which in a larger sample are so distributed as to make them equal, happen to be otherwise distributed· and if on the other hand the correlations really differ, this is still due to a sampling of the elemental factors

All this seems, to Thomson, to be merely a verbal expression of Pearson and Filon's formulae, and as long as the bonds causing correlation are positive, that is increase the variate in each instance if present, then there is bound to be some measure of hierarchical order among all correlation coefficients, the degree of perfection due to this cause being the point in dispute. Mr Udny Yule has declared that in his experience he has not found hierarchical order among coefficients, except in psychological measures. But it is very unusual to find, in other correlational fields, all the intercorrelations measured and set out, and though Thomson has unfortunately not had time from other duties to make a proper search, he finds some hierarchical order when the measurements permit of its discovery in the few instances examined. This point might well form a special inquiry.

Where the bonds are not all positive, but include interference factors which help one and hinder another variate, hierarchical order is probably less pronounced. If interference factors were as common as are mutually helpful factors, however, correlations would all tend to zero. It seems that in the realm of mental tests we have a province where the fact of general positive correlation implies that interference factors are in the minority. Positive bonds are the usual tendency. Whether these positive bonds are grouped entirely into a general factor *g*, leaving no residue, or an insignificant residue, of group factors, or are less uniquely grouped, appears to us still undetermined. But we must repeat that, while group factors may or may not be present, as long as the correlations are mainly positive a general factor may, of course, be postulated, and the controversy between us and Professor Spearman is not, and never was, as to the possibility of thus postulating a general factor, but as to the possibility of explaining all correlations thus without postulating any but the slightest group factors, and these very narrow in their action. Our position is that until the evidence is more clear we shall continue to suspect that numerous and wide group factors are present.

CHAPTER XII

THE MATHEMATICAL AND EXPERIMENTAL EVIDENCE FOR THE EXISTENCE OF A CENTRAL INTELLECTIVE FACTOR (g)*

[From *The British Journal of Psychology* (*General Section*), Vol XXIII, Part 2, October, 1932]

By WILLIAM BROWN

If a number of sufficiently dissimilar mental tests of intellective ability be applied to a group of individuals and correlation coefficients calculated, it is found that these correlation coefficients are related to one another in such a way that for any four (or *tetrad*) of them the following relation holds good, within the limits of random sampling, viz.

$$r_{ap}r_{bq} - r_{aq}r_{bp} = 0 \qquad \ldots\ldots (1),$$

and similarly with other arrangements of these four tests. We owe both the discovery of fact and the devising of the tetrad criterion to Professor C. Spearman.

The inference drawn from this is that the abilities measured by the mental tests are divisible into two factors each, the one being common to all (the general factor, g), while the other is in each case specific and independent (s).

In Professor Spearman's own words. "Whenever the tetrad equation holds throughout any table of correlations, and *only* when it does so, then every individual measurement of every ability (or of any other variable that enters into the table) can be divided into two independent parts which possess the following momentous properties. The one part has been called the 'general factor' and denoted by the letter 'g'; it is so named because, although varying freely from individual to individual, it remains the same for any one individual in respect of all the correlated abilities. The second part has been called the 'specific factor,' and denoted by the letter 's.' It not only varies from individual to individual, but even for any one individual from each ability to another†."

The relationship is expressed by the following equation:

$$m_{ax} = r_{ag}\ g_x + r_{as_a}\ s_{ax} \qquad \ldots\ldots (2),$$

* Communicated to Section J (Psychology) of the British Association for the Advancement of Science, London, Sept 25th, 1931—and, with certain additions, to the Tenth International Congress of Psychology at Copenhagen, Aug. 24th, 1932.

† C. Spearman, *The Abilities of Man*, London: Macmillan & Co Ltd. 1927, pp. 74, 75.

where m_{ax} = the measurement obtained for any individual x in the variable a, g_x = the individual's amount of g, the factor common to all the variables, and s_{ax} = the individual's amount of s_a, the factor specific to the variable a.

The method of applying the tetrad criterion is to draw up a frequency distribution of all the possible tetrad differences derivable from the table of correlation coefficients (there being $3\,{}^nC_4$ positive tetrad differences and an equal number of negative ones, where n is the number of mental tests correlated with one another) and to compare its standard deviation with the "theoretical" standard deviation of a purely chance distribution of such tetrad differences. A formula for the latter has been calculated by Spearman and Holzinger*, viz.

$$\bar{\sigma}_t^2 = \frac{4}{N}\left\{\bar{r}^2(1-\bar{r})^2 + \left[1 - 3\bar{r}\,\frac{(n-4)}{(n-2)} + 2\bar{r}^2\,\frac{(n-6)}{(n-2)}\right]s^2\right\} \quad (3),$$

where N = number of cases, n = number of tests, r = mean of correlation coefficients, s = standard deviation of correlation coefficients, and $\bar{\sigma}_t$ = the average value of the standard deviation of the tetrad differences.

In applying this formula, σ_t is generally multiplied by 0·67449 to give a "conventional" P.E., but since the frequency curve of tetrads is not, and cannot be, an exact probability curve (because the correlation coefficients, and therefore the tetrads, are not uncorrelated with one another), although it approximates to one, there is little to be gained by this procedure.

As the formula for the probable error of tetrads allows for the effects of "attenuation" upon correlation coefficients, so that these coefficients need not be corrected for observational errors, the "tetrad criterion" escapes criticism of mine† which was, as I still contend, rightly directed, in certain cases, towards the "correction formulae" which Professor Spearman devised in 1906 and 1910 to adjust for such errors. It also takes the place of the criterion of "intercolumnar correlation," or correlation between columns of correlations in a table of correlation coefficients, upon which Professor Spearman previously relied in proving his theory. According to this criterion, the average intercolumnar correlation should approximate to unity if the Theory of Two Factors holds good‡. But the correlation coefficients had first to be "corrected" for errors of observation (by formulae whose general or universal applicability I dispute), and even

* C. Spearman and K. Holzinger, "The Average Value for the Probable Error of Tetrad Differences," *Brit. Journ. Psychol.* 1930, XX, p. 370.

† *Essentials of Mental Measurement*, Cambridge: The University Press, 1st edition, 1911, pp. 83—85. *Brit. Journ. Psychol.* 1913, VI, p. 223.

‡ This is the famous "hierarchical order" of correlation coefficients.

then the criterion was only applicable to those coefficients which reached a certain "correctional standard," and thus did not admit of application to the whole table of correlation coefficients, as the tetrad criterion does.

In my own research work with mental tests in 1909, 1910 and 1913, I did not feel justified in considering my results as confirmatory of Professor Spearman's theory, because of the above-mentioned difficulties But on the other hand I never contended that my results *disproved* his theory In the *Essentials of Mental Measurement*, 1st edition, 1911, I wrote· "A definite solution of the question of the existence or non-existence of one central mental ability is yet to be sought It can only be obtained by the use of much larger random samples than those hitherto employed, since the probable errors must be small compared with the coefficients, if precise inferences are to be drawn from the latter, and in the case of small samples this condition is satisfied only for *large* correlation coefficients, which when obtained are often merely the result of selecting tests which measure closely similar mental abilities In all results hitherto quoted in support of ultimate identity of general intelligence and general sensory discrimination, the correlations contributed by the latter are so small compared with their P E 's that nothing definite can be inferred from them" (p. 120) My verdict at that date had to be "Non-proven," but certainly not "Disproved."

Nevertheless, in the interests of history and of scientific completeness, it seems worth while to work over some of those earlier results of mine with the aid of the tetrad criterion I have done this with (I) a group of 66 boys, aged 11–12 years, of an Elementary School (*Essentials of Mental Measurement*, 1st edition, p. 114), (II) a group of 40 boys, aged 11–12 years, of a Higher Grade School (*ibid* p. 116)*, and (III) a group of 83 boys, aged 14–16 years, of a Public School (St Paul's School), examined in mathematics only† In the first two groups I have pooled correlations between too-closely related abilities, in conformity with Professor Spearman's criticism, and in the third group I have refrained from "partialling out" for difference of form (the boys were drawn from five different forms) as this difference itself was some measure of difference of mathematical ability.

In Group I, with 8 tests, the total number of positive tetrad differences is 210. The observed median tetrad difference is 0·0208, and the esti-

* Both in "Some experimental results in the correlation of mental abilities," *Brit. Journ Psychol.* 1910, III. p 296.

† "An Objective Study of Mathematical Intelligence," *Biometrika*, 1910, VII. p. 352. The table of correlation coefficients, uncorrected for difference of form, is given in my *Mind and Personality*, University of London Press, Ltd. 1926, p. 123.

mated value is 0·024, which may be compared with the "theoretical" P E. value, calculated by the Spearman and Holzinger formula, 0·0353. Moreover, $\sigma_t = 0\ 0456$, which gives another value of the conventional P E, viz $0{\cdot}67449\sigma, = 0\ 03076$, which is also below the "theoretical" value These results, so far as they go, are all in favour of Spearman's two-factor theory. The distribution of tetrads is markedly leptokurtic, and $\beta_2 = 4{\cdot}275$, as compared with $\beta_2 = 3$ for a Probability Curve.

In Group II, with 8 tests and therefore 210 positive tetrads, the median tetrad difference is 0·0528 The "theoretical" P.E. (S and H) is 0 0495, which shows a difference of 0·0033 on the wrong side But $\sigma_t = 0{\cdot}069$, giving another value of the conventional P E. 0 0465, which is in conformity with the two-factor theory

The distribution of tetrads is markedly platykurtic $\beta_2 = 2{\cdot}358$.

In Group III, with 9 subsidiary mathematical abilities intercorrelated, and therefore 378 positive tetrads, the results are

$$\text{Median tetrad} = 0{\cdot}0761,\ \text{"Theoretical" P E (S. and H)} = 0{\cdot}0379,$$
$$\text{Mean tetrad} = 0{\cdot}1201,$$
$$\sigma_t = 0{\cdot}1612,\ \text{giving "Conventional" P E.} = 0{\cdot}1087,$$
$$\mu_2\ (\text{i.e. } \sigma_t^2) = 0\ 02601,\ \mu_4 = 0\ 002637 \quad \therefore\ \beta_2 = \frac{\mu_4}{\mu_2^2} = 3{\cdot}9041,$$
$$\text{i e } > 3 \text{ (a leptokurtic curve).}$$

Although this group furnishes a smooth frequency curve of tetrad differences, the results give little support to any theory of a central *mathematical* factor. No such theory has been put forward by Professor Spearman himself, and I have only introduced this group as a good illustration of the working out of the tetrad criterion on fairly adequate statistical material*. Whether the excess of σ_t over the corresponding Spearman-Holzinger value is statistically significant could only be decided with precision by determining the standard deviation of σ_t. This involves the use of complicated formulae devised by Professor Karl Pearson†.

But the results in Groups I and II, as far as numbers allow, do support the existence of a central intellective factor (g), and, when taken in relation with the large body of similar evidence accumulated during the last twenty years by Professor Spearman and his students, help to give

* My thanks are due to Mr R. J Bartlett, M Sc, for calculating tetrad differences and certain constants in this research I have myself re-calculated all the results, apart from the tetrad differences

† Applying these formulae, I find P E. of $\sigma_t = 0\ 001175$, and therefore P E of P E. $= 0{\cdot}0007928$. Hence the evidence is against the existence of a central *mathematical* factor.

solid basis to that theory. A mathematically satisfactory proof of the theory would involve much larger numbers, both of individuals tested and also of non-overlapping tests of intellective ability to be used upon them. In a critical article by Professor Karl Pearson* on this subject the following paragraph appears "We suggest that some 12 to 15 abilities (66 to 105 correlations, 1485 to 4095 tetrads), the abilities being settled by psychologists *a priori* to avoid 'overlaps,' are essential to a satisfactory test, the observations to be made on a homogeneous population of several hundreds. Short series involve such large probable errors that a mere statement that theory and observation are in accordance within the limits of the probable errors can carry no conviction with it"

I have organised a research along these lines, with the kind help of Professor Spearman and Dr W Stephenson of University College, London Dr Stephenson has devised a series of 20 tests of apparently non-overlapping intellective ability (selected not exactly *a priori* but after much careful preliminary trial), which received the approval of Professor Spearman, and has applied them for me to 300 boys, aged 10–10½ years, drawn from 12 Elementary Schools of the L C C , forming a homogeneous "random sample" of adequate size for statistical purposes The total number of positive tetrad differences is 14,535 It has since been found necessary to reject one of the tests and one of the correlation coefficients There remain 11,356 positive tetrads which form a smooth frequency curve, the mathematical properties of which I am now working out† I have found that the best-fitting frequency curve is a Type II A Pearson curve, with equation

$$y = 1412\left(1 - \frac{x^2}{1188}\right)^{13\cdot 669}. \quad [\text{Unit of grouping} = 0\cdot 005\,]$$

The curve is platykurtic, with $\beta_2 = 2{\cdot}81446$ The standard deviation, σ_t, $= 0{\cdot}031289 \pm 0{\cdot}002586$ If we compare this with the "theoretical" Spearman-Holzinger value, $\bar{\sigma}_t = 0{\cdot}02827$, we find an excess of 0·003019, being 1·167 times the probable error This indicates a good correspondence of observation with theory. But only after much further psychological and mathematical analysis of the enormous mass of data can a final conclusion be drawn. The material satisfies the most stringent demands of statistical theory and will furnish a precise and definite solution of the problem of a Central Intellective Factor.

* Karl Pearson and Margaret Moul, "The mathematics of intelligence, I The sampling errors in the theory of a generalised factor," *Biometrika*, 1927, XIX p. 261

† This is set out in the following chapter (Chapter XIII)

APPENDIX

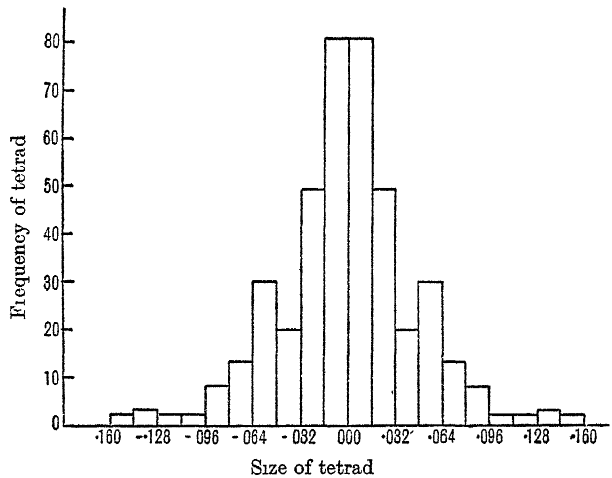
GROUP I. 66 *boys, aged* 11–12 *years. Elementary School*

Correlation Square

	1	2	3	4	5	6	7	8
1 Combination	—	0 52	0 52	0 39	0 46	0 13	0 00	0 15
2 Memory, poetry	0 52	—	0 49	0 39	0 27	0 12	0 05	0 13
3 Memory, mechanical	0 52	0 49	—	0 29	0 34	0 14	0 12	0 10
4. Addition	0 39	0 39	0 29	—	0 41	0 12	0 03	0 20
5 Letters (ER + ANOS)	0 46	0 27	0 34	0 41	—	0 37	0 10	0 00
6. Motor Ability	0 13	0 12	0 14	0 12	0 37	—	0 04	0 00
7 Illusion (M -L)	0 00	0 05	0 12	0 03	0 10	0 04	—	0 16
8. Bisection	0 15	0 13	0 10	0 20	0 00	0 00	0 16	—

No of +ve tetrad differences of form $r_{13}\, r_{24} - r_{14}\, r_{23} = 3 \times {}^8C_4 = 210$

Symmetrical Distribution of Tetrad Differences



Distribution of Positive Tetrads

Range of tetrad	0 000–0 016	0 016–0 032	0 032–0 048	0 048–0 064	0 064–0 080	0 080–0 096	0 096–0·112	0 112–0 128	0 128–0 144	0 144–0 160	Total
Frequency	81	49	20	30	13	8	2	2	3	2	210

Observed median tetrad difference = 0 0208.
Estimated " " " = 0 024.
Observed σ of tetrad differences = 0 0456
. Conventional P E. = 0·67449σ = 0 03076
"Theoretical" P E (Spearman and Holzinger) = 0 0353.
$\mu_2 = 0\ 00207846$, $\mu_4 = 0\ 000018469$,

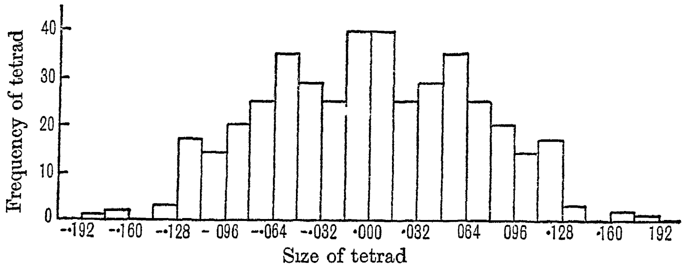
$$\therefore\ \beta_2 = \frac{\mu_4}{\mu_2{}^2} = 4\ 275 \text{ (a leptokurtic curve).}$$

Group II. 40 *boys, aged* 11–12 *years Higher Grade School*

Correlation Square

		1	2	3	4	5	6	7	8
1	Intelligence (S Marks + Gen Intell.)	—	0·585	0·645	0·465	0·26	0·23	0·445	0·275
2	Memory, poetry	0·585	—	0·44	0·44	0·00	0·185	0·38	0·19
3	Combination	0·645	0·44	—	0·46	0·32	0·05	0·28	0·28
4	Drawing	0·465	0·44	0·46	—	0·14	0·15	0·39	0·00
5	Addition (speed)	0·26	0·00	0·32	0·14	—	0·275	0·00	0·20
6	Letters (ER + ANOS)	0·23	0·185	0·05	0·15	0·275	—	0·00	0·125
7	Memory, mechanical	0·445	0·38	0·28	0·39	0·00	0·00	—	0·00
8	Motor Ability (all letters)	0·275	0·19	0·28	0·00	0·20	0·125	0·00	—

Symmetrical Distribution of Tetrad Differences



Distribution of Positive Tetrads

Range of tetrad	0·000–0·016	0·016–0·032	0·032–0·048	0·048–0·064	0·064–0·080	0·080–0·096	0·096–0·112	0·112–0·128	0·128–0·144	0·144–0·160	0·160–0·176	0·176–0·192	Total
Frequency	39	25	29	35	25	20	14	17	3	0	2	1	210

Median tetrad difference = 0·0528

Observed σ of tetrad differences = 0·069.

∴ Conventional P.E. = 0·0465.

"Theoretical" P.E. (S. and H.) = 0·0495.

$\mu_2 = 0{\cdot}004755$, $\mu_4 = 0{\cdot}000053308$,

∴ $\beta_2 = 2{\cdot}358$ (a platykurtic curve).

Group III. 83 *boys, aged* 14–16 *years Public School*

Mathematical Examination, 1910

Correlation Square

	C	H	E	I	G	A	F	D	B
C	—	0·57	0·47	0·55	0·59	0·78	0·51	0·81	0·60
H	0·57	—	0·69	0·92	0·53	0·61	0·55	0·40	0·26
E	0·47	0·69	—	0·57	0·76	0·44	0·82	0·46	0·23
I	0·55	0·92	0·57	—	0·49	0·47	0·49	0·43	0·22
G	0·59	0·53	0·76	0·49	—	0·41	0·64	0·44	0·26
A	0·78	0·61	0·44	0·47	0·41	—	0·46	0·65	0·49
F	0·51	0·55	0·82	0·49	0·64	0·46	—	0·49	0·28
D	0·81	0·40	0·46	0·43	0·44	0·65	0·49	—	0·37
B	0·60	0·26	0·23	0·22	0·26	0·49	0·28	0·37	—

Geometry

A Memory of Definitions and General Principles (e g principle of superposition).
B. Memory of constructions
C. Memory of preceding propositions and power of applying them
D. Recognition of necessity of generality in proof, and power of recognising general relations in a particular case

Arithmetic

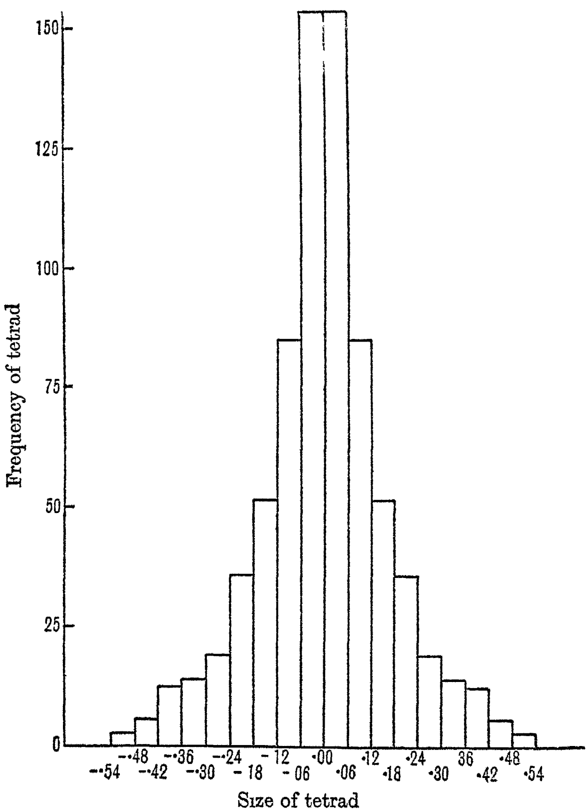
E Accuracy.
F General memory of rules and power of applying them.
G Power of doing sums in percentage and proportion

Algebra

H Accuracy.
I General memory of rules and power of applying them.

No of +ve tetrad differences of form $r_{13}r_{24}-r_{14}r_{23}=3\times {}^9C_4=378$.

Symmetrical Distribution of Tetrad Differences



Distribution of Positive Tetrads

Range of tetrad	0·00–0 06	0 06–0 12	0 12–0 18	0 18–0 24	0·24–0 30	0 30–0 36	0 36–0 42	0 42–0 48	0 48–0 54	Total
Frequency	153	85	52	36	19	13	12	6	2	378

The constants of this distribution are

Median tetrad $=0\ 0761$. Standard deviation, σ_t, $=0\ 161276$.

Mean tetrad $=0\ 1201$. $0\ 67449\sigma_t=0\ 10878$.

"Theoretical" P E. (S. and H) $=0\ 0379$

$\sigma_t^2=\mu_2=0\ 02601$, $\mu_4=0\ 002637$

$$\therefore\ \beta_2=\frac{\mu_4}{\mu_2^2}=3\ 9041 \text{ (a leptokurtic curve).}$$

The best-fitting curves to the distributions are Type II A (platykurtic) and Type II B (leptokurtic) Pearson curves The actual equations to the three curves are

for Group I, $$y=\frac{65\ 11}{\left(1+\frac{x^2}{54\ 445}\right)^{4\ 853}},$$

for Group II, $$y=34\ 754\left(1-\frac{x^2}{136\ 43}\right)^{2\ 174};$$

for Group III, $$y=\frac{121\ 63}{\left(1+\frac{x^2}{62\ 3976}\right)^{5\ 818}}.$$

CHAPTER XIII

A TEST OF THE THEORY OF TWO FACTORS

[From *The British Journal of Psychology (General Section)*, Vol. XXIII, Part 4, April, 1933]

By WILLIAM BROWN AND WILLIAM STEPHENSON

(1) INTRODUCTION

IN a critical article by Professor Karl Pearson (1) on the Theory of Two Factors it is suggested that "some 12 to 15 abilities   the abilities being settled by psychologists *a priori* to avoid 'overlaps,' are essential to a satisfactory test [of this theory], the observations to be made on a homogeneous population of several hundreds" (p 261) The present paper describes a test of the kind that Professor Pearson demands, using correlations for some 20 abilities The research was organised by one of us (W. Brown) over a year ago, in connection with his re-testing of his own earlier correlational material by Professor Spearman's "tetrad criterion" (2), and he has made himself particularly responsible for the mathematical arguments and conclusions of the investigation, including the curve-fitting. But the mental tests were devised and applied by Dr W. Stephenson, and the correlation coefficients and tetrad differences were also calculated by him Some preliminary remarks, from both psychological and mathematical viewpoints, are called for.

(2) THE VIEWPOINT OF EXPERIMENTAL PSYCHOLOGY

The psychologist does not select the 20 abilities on a narrow *a priori* basis. The abilities are found by a slow process of experimentation and test refining. Moreover, this experimentation is itself based upon the theory of two factors. The psychologist devises tests that, approximately, should fit a theoretical criterion (that of zero "tetrads")*.

But it betrays a misconception of the nature of a scientific theory to say that, thereby, the psychologist is working in a closed circle. On the contrary, he works like a physicist; he establishes the fact that under certain conditions the criterion is satisfied, and he determines the nature of these conditions. The core of the matter is that in this way, and generally, the theory of two factors *works* for psychology. When reason-

* Other tests can be devised, likewise, that should *not* fit the criterion, as is the case when tests are too similar.

able agreement is found between the criterion and correlational facts, the common factor can receive an acceptable psychological explanation When the criterion and facts do not agree the psychologist makes a determination of "overlap," "group factors," or "specificality*." Again (and this is the essential matter), it is found that either the specificalities have acceptable psychological explanations, the true influence of which had been neglected, or that a field of research is opened up for facts that are essentially new and unexpected

Thus, what for mathematics is a failure of the criterion is for psychology a pointer to new psychological findings The proof of the theory is much more than a purely mathematical matter There is in the proof the foundation and development of a scientific experimental psychology; and, although we would be modest, to that extent it constitutes a "Copernican revolution" The mathematician, given the data, might prove the theory to acceptable limits—it is so proved, even for the exacting conditions laid down by Professor Pearson, in the course of the present paper. But only the psychologist can now disprove the theory

The mathematics and the psychology of the theory of two factors are now so developed that the mathematician, before he can test out some of his sub-theories, can scarcely proceed unless the experimental data are as precisely as possible of the kind that the psychologist daily tries to supply for the progress of his science. It seems that neither the mathematics nor the psychology can be absolutely rigorous, any more than is the case in physics, and neither can proceed without the other.

(3) AN EXAMPLE OF THE WORKING OF THE THEORY

An illustration of the *modus operandi* of the theory of two factors can serve to amplify the above viewpoint and introduce some of the requirements of a test of the theory.

It was found by one of us (3) that verbal intelligence tests did not, whereas certain non-verbal tests did, agree satisfactorily with the tetrad criterion. A theory of an additional *v*-factor, a group factor, was tried out for the verbal tests, and with further researches the full force of the influence of reproduction and experience on these verbal tests was made apparent. It now appears that only when past achievement, and therefore the rôle of retentivity and reproduction, is rendered as simple as possible, as when primarily perceptual tests are used, does the criterion agree well with the facts. A rôle for past experience and reproductive processes can scarcely be denied for the verbal tests, and the theory of

* The term "specificality" covers both positive ("overlap," or "group factor") and subtractive or negative influences. It stands for any disturbance of the tetrad criterion.

two factors *works* because it isolates in a v-factor the activities in which experience and reproduction are strongest and most to be expected. *Pari passu*, the theory works in so isolating a purer field of perceptual tests within which eductive processes can be considered to hold sway.

(4) REQUIREMENTS FOR A TEST OF THE THEORY

The important requirement is the set of 20 or so non-overlapping abilities.

It has been shown (4) that non-verbal tests, that were primarily perceptual in form, supplied correlations agreeing with the theory of two factors for a population of 1000, and, as has been suggested above, the non-overlapping tests are most likely to be those of perceptual foundation. The work of Line (5) must be mentioned as a pioneer study with such a test. At the time of beginning the present research there were not available 20 sufficiently developed primarily perceptual tests But some non-verbal tests can be used, with the knowledge that many serve as "pure" g-tests, each with g-factor and factors specific to the test, the latter condition obtaining because of the narrow range of the experiential or reproductive influences in them For this reason two or more non-verbal tests, themselves not perceptual in any critical way, might act as "reference values" for the primarily perceptual tests. The common visual fundaments of the latter may thus be controlled for specificality, although it seems impossible to conceive of any test that is less critical for all but eductive processes than these latter We had available, then, 11 primarily perceptual, and 5 non-verbal, tests.

Now any *one* verbal test could be used together with the above 16 tests, in spite of v-factor content, because the v-factor would then be specific relative to the set-up of tests. But, in spite of their v-factor content, there were compensating reasons that made it desirable to include more than one verbal test in our battery. Verbal tests lend themselves to a freshness of working that helps along a long day's testing, and they introduce high correlations with some of the perceptual and non-verbal tests. Of greater moment, there would seem to be no reason why a partialling device should not be used, whereby the v-factor is partialled out, leaving the intercorrelations amongst the verbal tests attributable to a g-factor, so that these partialled correlations might then be used as non-overlapping values, then comparable with the correlations amongst the perceptual and non-verbal tests themselves. Evidence could be sought in the present work for the applicability of such a partialling technique. Thus, we added 6 verbal tests to our battery, making 22 tests in all.

With these tests specially prepared for group application to children of the age upon which it was decided to work, we have only to guard against sources of specificality of the varieties described by Spearman (6), and data should ensue that are suitable for the mathematician's detailed examination. A test of the theory of two factors requires a correlation table showing steep hierarchy, so that some of the tests need not be highly intercorrelatable. It would seem that a population of 300 is a minimum for use in fine correlational work

(5) THE TESTS AND THEIR APPLICATION

Table I names the various tests, in order of application to the testees The following notes will help to give meaning to the table.

Three hundred boys were tested in groups of not more than 25 per group. The boys were of age 10 to 10½ years at the time of testing, and

Table I. *Showing the tests in order of application*

Testing period	No.	Type	Test name	Time allowed for demonstration min.	Time allowed for the test proper min
1st	1	*v*	Inventive synonyms	2	3
	2	*n*	Alphabetical form	5	4
	3	*n*	Alphabetical series	3	8
	4	*v*	Disarranged sentences	2	6
	5	*p*	Fitting shapes	3	8
	6	*v*	Understanding paragraphs	1	12
2nd	7	*p*	Mazes	2	4
	8	*n*	Cancellation	3	3
	9	*p*	Pattern perception	4	5
	10	*p*	Analogies form	5	6
	11	*p*	Classification "rights"	6	12
3rd	12	*n*	Mutilated pictures	2	4
	13	*p*	Overlapping shapes	6	6
	14	*v*	Inferences, selective	2	6
	15	*p*	Abstraction "pairs"	4	8
	16	*p*	Code	2	3
	17	*p*	Code-parts	2	3½
4th	18	*v*	Classification, selective	2	4
	19	*n*	Arithmetical equations	4	6
	20	*v*	Proverbs, selective	2	10
	21	*p*	Series form	5	8
	22	*p*	Pitch perception	5	10

were all the boys of that age in the two or three elementary school classes usually accommodating that age. Boys in lower classes, or those who suffered obviously from physical or scholastic disabilities, were not tested. Each group of boys attempted all 22 tests on the one school day, in four testing periods of about an hour each, commencing at 9.30 a.m., and ending at 4 p.m. A standard procedure was followed of demonstrating

at least six sample test-units on the school blackboard just before beginning each test, and then allowing a fore-practice period to the testees, using test-units printed on the covering page of the test proper. The time allowance for demonstration and fore-practice, and for the test proper, is shown in Table I for each test.

A number is given to each test in Table I: the letter "*p*" following the test number indicates that the test is primarily perceptual, whilst "*n*" and "*v*" indicate that the tests are non-verbal and verbal respectively.

The tests numbered 1, 2, 3, 4, 5, 6, 10, 12, 13, 18 are described in previous papers (4); and in each case the test used in the present work was a development of that of the same name used in the work on 1000 children. Tests 10 and 11 are of the types developed by Fortes (7) and Line (5) respectively. No. 22 is the "Pitch Perception" of Seashore's Musical Ability Test (8), applied by gramophone. No. 7 consisted of a set

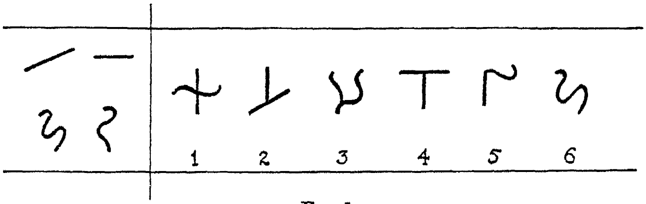


Fig. 1

of mazes of the well-known Porteus kind, but here prepared specially for group application. Nos. 8, 14 and 20 are considered to be sufficiently known by name. It is not considered necessary to enter into further descriptive details about any of the tests mentioned above, so that only the tests numbered 15, 16, 17, 9 and 19 need be given a short description.

Fig. 1 shows a test-unit of the "Abstraction" test (No. 15). The pair of drawings at the left-top are *alike* in a certain way, the second pair at the left-bottom are also *alike* in a certain way; and two of the six drawings numbered 1 to 6 have to be selected in which *both* the above-determined likenesses are to be found.

The Code test (No 16) is a Substitution test. In No. 17 the same code items are used, but only *parts* of the code items are now provided, and the testee is required to substitute under each part the code number of the complete item. Fig 2 (*a*) shows the code items used in test No. 16; and Fig. 2 (*b*) shows a short row of code-parts of test No. 17—the numbers 3, 2, 4, 1, 5, 6, .. etc., have to be substituted under these code-parts.

In test No. 9 (Pattern Perception) a pattern is given at the left, and this can be found exactly in the more complicated pattern at the right.

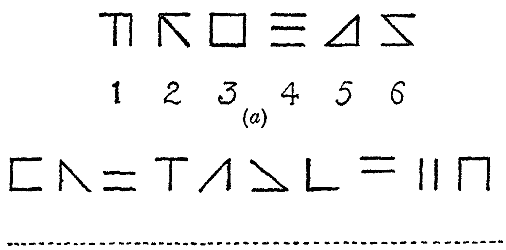
a line has to be drawn around the determined pattern at the right. A test-unit is shown in Fig 3

A sample test-unit of the Arithmetical Equations test, No. 19, is as follows:

$$3 \quad 9 \quad 5 = 7.$$

The signs + and − have to be placed between the numbers at the left side of the equation, so that the left side then equals the right side, i e.

$$3 + 9 - 5 = 7.$$



(b)

Fig. 2

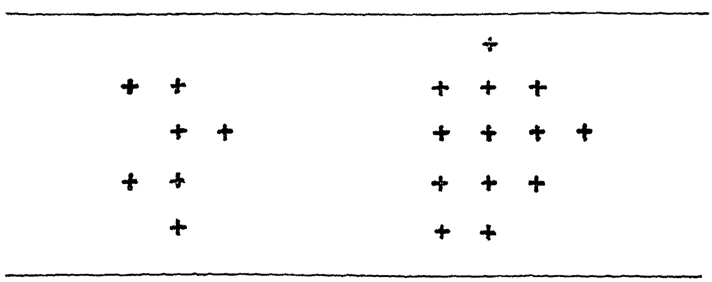


Fig. 3

(6) SCORING AND CORRELATION CALCULATION

The tests were scored by two competent psychologists, one test at a time for any one group of papers.

Because almost all the tests had been used previously in work on 10–11 year old groups, and were developed in accordance with results obtained for this age, the scatter of the score in each test was approximately "*normal.*" But, to obviate any disturbances in tetrads attributable to non-normal score distributions, a first step in the calculational work consisted in converting the crude scores for each test to values fitting a standard scale. The standard scale, used for every test, ranged from

score 0 to score 19, with frequencies 1, 2, 4, 7, 10, 15, 21, 27, 30, 33, 33, . . . and again downwards to score 1, for the scores 0 to 19 respectively*.

Correlations were calculated, using the "difference" method† The "differences" were squared and added, using Burrough's Adding Machine with tape recording Seven-figure logarithms were used in all calculational work. But the available multiplication tables that were used later were for tetrad calculations for three figures only, so that it is sufficient to report correlations to three places of decimals only.

(7) CORRELATIONS

Table II shows the first correlation table for 20 of the 22 tests The tests numbered 21 and 22 are not used in the present work, because 20 is a sufficiently large number for our present purpose. The two, 21 and 22, were held in reserve, they were the tests last applied in the day, and this is the only reason why they, rather than any other of the 22 tests, are held in reserve.

But Table II cannot be employed as it stands for the test of the two-factor theory. Six of the tests are verbal, entailing a v-factor, and two of the non-verbal, perceptual tests also entail specificality. There is required a procedure for partialling out such specificalities, and the following was adopted

(i) First the g-factor is partialled out of the set of tests involving a known specificality, using perceptual tests as "reference values" Thus, taking the v-factor as a known specificality, the g-saturation of any verbal test is given by the following equation‡:

$$r^2_{vg} = \frac{r_{vp_1} \cdot r_{vp_2} + r_{vp_1}\ r_{vp_3} + \cdots}{r_{p_1p_2} + r_{p_1p_3} +}.$$

Using g-saturation values obtained in this way, the g-factor is partialled out of the tests involving the known specificality by use of the following equation:

$$r_{v_1v_2 \cdot g} = \frac{r_{v_1v_2} - r_{v_1g}\ r_{v_2g}}{\sqrt{1-r^2_{v_1g}}\sqrt{1-r^2_{v_2g}}}.$$

(ii) Using the g-partialled correlations ("specific" correlations) given above, the specificality saturation is next determined. This repeats the first step of (i) above: thus, the v-saturation of any verbal test is given in terms of the other verbal tests by the following equation

$$r^2_{v_1g\ p} = \frac{r_{v_1v_2\ g} \cdot r_{v_1v_3\ g} + \cdots}{r_{v_2v_3\ g} +}.$$

* For a description of the converting method, see Stephenson(4).

† $$r = \frac{\Sigma X^2 + \Sigma Y^2 - \Sigma (X-Y)^2 - 2nm_x m_y}{2\sqrt{\Sigma X^2 - nm_x^2}\ \sqrt{\Sigma Y^2 - nm_y^2}}$$

‡ The subscripts "v" and "p" refer to "verbal" and "perceptual" tests respectively. The notation used in these first sections must not be confused with that of the Statistical Evaluation of Results, p 220.

(iii) Now restart with the first correlations, and partial out the specificality, using the specificality-saturations given at (ii) Thus, to partial out the v-factor use the v-saturations determined as at (ii) in the following equations.

$$r_{v_1 v_2 \cdot v} = \frac{r_{v_1 v_2} - r_{v_{1g} \cdot v}\, r_{v_{2g} \cdot v}}{\sqrt{1-r^2_{v_{1g} \cdot v}}\,\sqrt{1-r^2_{v_{2g} \cdot v}}}.$$

The step (ii) can only be taken when the specificality correlations given at (i) themselves fit the tetrad criterion. there must therefore be at least four tests involving the same specificality, although three tests serve to determine specificality saturations. In most cases, however (other than for well-defined group factors like the v-factor), specificality is observed for *two* tests only, but to a first approximation the square root of the specific correlation can be taken to be the specificality saturation of each of the two tests Using this saturation value, the specificality can be partialled out as at step (iii) above

The above partialling procedures are theoretically warranted in our work, and we do not offer any further substantiation of them, but it should be observed that the probable errors of the above partial correlations are not known, although they can be considered to be a little greater than those for the first correlations.

Apart from the v-factor there is obvious specificality in two correlations only in Table II, for the two Code tests (Nos. 16 and 17), and for two perceptual tests (Nos. 5 and 9). The Code tests are, of course, too similar to be free from overlapping. Using the above partialling procedure we can partial out Code-specificality, leaving a correlation $r_{16\ 17}$ attributable to g-factor only. In this way the first correlation 0·644 is reduced to 0·452. As for the tests 5 and 9, it was the first time that both had been used together in a correlation table, and the tests are apparently too similar in some way, the spatial fundaments are similar and both require a certain drawing ability. Both have fairly easy test-units, and thus "speed-preference" is not an unlikely disturbance. But the correlation $r_{5\ 9}$ need not concern us further, we shall find later that the test No. 5 is the source of other disturbances The above partialling procedure results in a value 0 518 in place of 0·655 for $r_{5\ 9}$, the value 0·518 now being attributable to g-factor only.

A correlation table "corrected" for v-factor, Code-specificality, and specificality between the two similar perceptual tests, is given as Table III, the correlations now being reported in hierarchical order They are now ready for tetrad examination. But, again, only after the tetrad examination can we decide whether the correlations are acceptable as "non-overlapping" values.

Table II. *Showing first correlations for 20 tests, population of 300 boys, age* $10\text{–}10\frac{1}{2}$ *years*

(The numbers at the extreme left of the Table refer to the corresponding numbers in Table I)

	Inventive synonyms	Disarranged sentences	Understanding paragraphs	Inferences	Classification (verbal)	Proverbs	Alphabetical form	Alphabetical series	Cancellation	Mutilated pictures	Arithmetical equations	Fitting shapes	Mazes	Pattern perception	Analogies	Classification "rights"	Overlapping shapes	Abstraction "pairs"	Code	Code-parts
1. Inventive synonyms	—	656	764	589	657	577	451	494	274	345	489	373	305	429	374	448	373	408	352	403
4 Disarranged sentences		—	702	566	659	500	442	602	344	396	584	465	324	516	507	481	467	499	436	496
6 Understanding paragraphs			—	644	645	620	399	571	363	358	585	453	372	525	476	520	465	507	456	488
14 Inferences				—	625	548	385	546	250	300	521	374	296	426	454	486	419	435	415	469
18. Classification (verbal)					—	521	420	526	248	285	532	404	268	414	447	449	389	409	388	440
20. Proverbs						—	327	418	174	246	446	315	203	340	421	417	315	386	324	388
2. Alphabetical form							—	515	293	277	501	476	300	411	417	416	398	359	304	474
3. Alphabetical series								—	332	322	549	570	407	534	575	593	508	477	440	491
8 Cancellation									—	176	384	312	314	364	359	376	277	257	443	408
12 Mutilated pictures										—	350	394	334	401	398	309	266	289	245	375
19 Arithmetical equations											—	425	358	549	522	513	491	460	444	557
5 Fitting shapes												—	429	655	542	475	437	359	345	510
7 Mazes													—	404	407	318	267	285	322	442
9 Pattern perception														—	541	468	441	394	391	547
10. Analogies															—	506	441	463	397	465
11 Classification "rights"																—	468	477	393	432
13 Overlapping shapes																	—	435	459	436
15 Abstraction "pairs"																		—	453	482
16 Code																			—	644
17 Code-parts																				—

Table III. *Showing correlations for 20 tests in hierarchical order, v-factor and two specificalities partialled out*

(The numbers at the extreme left refer to the same numbers in Tables I and II)

			Alphabetical series	Arithmetical equations	Understanding paragraphs	Disarranged sentences	Code-parts	Analogies	Pattern perception	Classification "rights"	Inferences	Fitting shapes	Classification (verbal)	Abstraction "pairs"	Inventive synonyms	Overlapping shapes	Alphabetical form	Code	Proverbs	Mazes	Mutilated pictures	Cancellation
			a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*	*q*	*r*	*s*	*t*
3	Alphabetical series	*a*	—	549	571	602	491	575	534	593	546	570	526	477	494	508	515	440	418	407	322	332
19	Arithmetical equations	*b*		—	585	584	557	522	549	513	521	425	532	460	489	491	501	444	446	358	350	384
6.	Understanding paragraphs	*c*			—	558	488	476	525	520	474	453	368	507	500	465	399	456	445	372	358	363
4	Disarranged sentences	*d*				—	496	507	516	481	418	465	510	499	481	467	442	436	333	324	396	344
17.	Code-parts	*e*					—	465	547	432	469	510	440	482	403	436	474	452	388	442	375	408
10	Analogies	*f*						—	541	506	454	542	447	463	374	441	417	397	421	407	398	359
9.	Pattern perception	*g*							—	468	426	518	414	394	429	441	411	391	340	404	401	364
11.	Classification "rights"	*h*								—	486	475	449	477	448	468	416	393	417	318	309	376
14.	Inferences	*i*									—	374	467	435	385	419	385	415	405	296	300	250
5.	Fitting shapes	*j*										—	404	359	373	437	462	345	315	429	394	312
18.	Classification (verbal)	*k*											—	409	392	389	420	388	324	268	286	248
15.	Abstraction "pairs"	*l*												—	408	435	359	453	386	285	289	257
1.	Inventive synonyms	*m*													—	373	416	352	374	305	345	274
13	Overlapping shapes	*n*														—	398	408	315	267	266	277
2.	Alphabetical form	*o*															—	304	327	300	277	293
16.	Code	*p*																—	324	322	245	343
20	Proverbs	*q*																	—	203	246	173
7.	Mazes	*r*																		—	334	314
12.	Mutilated pictures	*s*																			—	176
8.	Cancellation	*t*																				—

(8) THE PROBLEM OF FURTHER SPECIFICALITIES

There are 14,535 tetrad differences for a table of 20 tests These were set down by Stephenson, and products and differences were computed by a well-trained and capable worker, using Cotsworth's "Direct Calculator"

Only some 23 of the 14,535 tetrad differences have values greater than 0·1000. An examination of the tetrads shows that the correlation $r_{6\ 18}$ (for the two verbal tests, "Understanding Paragraphs" and "Classification") accounts for seven of these large values, and that the mean of all tetrads involving this correlation is more than five times probable error*. This is a partial correlation, and its probable error is likely to be larger than that for a first correlation but, even so, the two tests are at the opposite extremes of quality-quantity preference—the "Classification" is a *speed* test of a marked kind (only four words have to be regarded in each test-unit, one of which is *unlike* the other three), whilst the "Understanding Paragraphs" is much more a *power* test (involving the regard of a long paragraph, followed by answering questions about the paragraph) "Speed preference" gives an advantage to the former, and a disadvantage to the latter test, and a misbalance of this kind frequently leads to specificality for the two tests concerned. It thus seems allowable to discard tetrads which involve this correlation $r_{6\ 18}$. (We could, if need be, re-score the two tests, penalising for "speed" in the one case, and allowing extra marks for it in the other, and so remove the disturbance)

Finally, one test, that of "Fitting Shapes," can be considered to account for nearly all other tetrad differences greater than 0·1000, the correlations of this test with tests numbered 2, 7, 9, 10 and 17 being particularly concerned. It is of interest that these tests amongst themselves appear to show some slight signs of a group factor, on the borderline of significance All these tests involve "spatiality," and either this or speed preference could account for a group factor. If test No. 5 ("Fitting Shapes") was the most highly saturated of a set of tests involving such a group factor, then the data we have obtained are explicable. It would seem acceptable to omit the test No. 5 from tetrad consideration. At worst it is no sin to omit *one* test from a battery of so many, even were there no readily acceptable explanation to be offered to the specificality that it entails

After removing the test No. 5, and all tetrads involving the correlation $r_{6\ 18}$ (and these only), there remain 11,356 positive tetrad differences,

* Using an approximate probable error, given by C. Spearman, *Abilities of Man*, Appendix, p x, formula 14

with frequency distribution given in Fig. 4 This is the first set of tetrads that are offered as those obtained for non-overlapping tests and that can receive the statistician's attention

(9) STATISTICAL EVALUATION OF RESULTS

We have first to determine the constants of the distribution of the 170 correlation coefficients Grouping in intervals of 0·05, we have

$$\bar{r} = 0{\cdot}41353 \text{ (0·41367 by actual averaging of separate } r\text{'s)},$$
$$\sigma_r = \sqrt{\mu_2} = 0{\cdot}087268;$$
$$\mu_2 = 0{\cdot}0076157; \quad \mu_3 = -0{\cdot}00016834, \quad \mu_4 = 0{\cdot}00015647,$$
$$\beta_1 = 0\ 064157; \quad \beta_2 = 2\ 6978, \quad \therefore\ \beta_4 = 11 \text{ (approx)*}.$$

(These are the values after Sheppard's corrections had been applied The uncorrected values were

$$\sigma_r = 0{\cdot}088398, \quad \mu_2 = 0{\cdot}007824; \quad \mu_3 = -\ 0\ 00016834,$$
$$\mu_4 = 0{\cdot}00016623; \quad \beta_1 = 0{\cdot}05917, \quad \beta_2 = 2{\cdot}7219\ )$$

Accordingly from the Spearman-Holzinger formula (9)

$$\bar{\sigma}_t{}^2 = \frac{4}{N}\left\{\bar{r}^2(1-\bar{r})^2 + \left[1 - 3\bar{r}\frac{(n-4)}{(n-2)} + 2\bar{r}^2\frac{(n-6)}{(n-2)}\right]\sigma_r{}^2\right\},$$

where N = no. of cases (300) and n = no. of tests (19)

$$= 0{\cdot}00079937,$$
$$\bar{\sigma}_t = 0{\cdot}02827.$$

[Taking $\bar{r} = 0{\cdot}41367$, $\bar{\sigma}_t = 0{\cdot}0282612$.]

In order to be able to compare this "theoretical" value with our "observed" result (to be given later) we must determine its "probable error." For this purpose we have employed the following formulae, given in the article of Karl Pearson and Margaret Moul (1)†:

$$\sigma^2_{\bar{r}} = \frac{1}{Np}\left[(1-\bar{r}^2)^2 - 2\mu_2(1-3\bar{r}^2) + 4\bar{r}\mu_3 + \mu_4 + (n-2)\bar{r}(1-\bar{r})^2(2+n\bar{r})\right]$$
$$= 0{\cdot}0468664 \ (p = \text{no. of correlation coefficients} = 170),$$
$$\sigma^2_{\mu_2} = \frac{4}{Np}\left[(1-\bar{r}^2)^2\mu_2 - 4\bar{r}(1-\bar{r}^2)\mu_3 + 2(1+\bar{r}^2)\mu_4 - 4\bar{r}\beta_4\mu_2{}^3\right]$$
$$(\beta_4 = 11, \text{ approx.})$$
$$= 0{\cdot}00000044839;$$

* *Tables for Statisticians and Biometricians*, vol. I, Table XLII (*b*), p 78.

† Pp. 258, 259, 260. We have written p = no of correlation coefficients, and n = no of tests, reversing Pearson's notation, for the sake of uniformity with the Spearman-Holzinger formula.

$$\{\delta\bar{r}\,\delta\mu_2\} = \frac{2}{Np}\left[-4\bar{r}(1-\bar{r}^2)\mu_2 - 2(1-3\bar{r}^2)\mu_3 + 4\bar{r}\mu_4\right]$$
$$= -0{\cdot}00000039295,$$

$$\sigma^2_{\sigma_t^2} = \frac{64}{N^2}\left(\bar{r}(1-\bar{r})(1-2\bar{r}) + \frac{3(1-2\bar{r})}{p-1}\mu_2\right)^2\sigma^2_{\bar{r}}$$
$$+\frac{16}{N^2}\left(1+\frac{6\bar{r}(1-\bar{r})}{p-1}\right)^2\sigma^2_{\mu_2}$$
$$+\frac{64}{N^2}\left(\bar{r}(1-\bar{r})(1-2\bar{r}) + \frac{3(1-2\bar{r})}{p-1}\mu_2\right)\left(1+\frac{6\bar{r}(1-\bar{r})}{p-1}\right)\{\delta\bar{r}\,\delta\mu_2\}$$
$$= 0{\cdot}000000057591106;$$

$\therefore\ \sigma_{\sigma_t^2} = 0{\cdot}00023998,$

$\sigma_{\sigma_t} = \sigma_{\sigma_t^2}/2\sigma_t$, assuming approx. normal distribution ((1), p. 268)

$= 0{\cdot}0038349;$

$\therefore$ P.E. of $\bar{\sigma}_t = \mathit{0{\cdot}0025866}$.

Thus the "theoretical" value, $\sigma_t = \mathit{0{\cdot}02827} \pm \mathit{0{\cdot}0025866}$.

Turning now to the observed frequency distribution of tetrads, grouped in subranges of 0·005, we have the following distribution constants:

Median (Quartile of Symmetrical Distribution) = 0·0214,

$$\sigma_t = 0{\cdot}031289, \qquad \mu_2 = \sigma_t^2 = 0{\cdot}000979, \qquad \mu_4 = 0{\cdot}00000269749;$$

$$\therefore\ \beta_2 = \frac{\mu_4}{\mu_2^2} = 2{\cdot}81446 \text{ (a platykurtic curve).}$$

As the tetrad differences must lie numerically between -1 and $+1$, the best-fitting curve must be a limited range symmetrical curve. Since $\beta_2 < 3$, the best-fitting curve is a Type II A Pearson curve*, with equation

$$y = y_0\left(1-\frac{x^2}{a^2}\right)^m.$$

Here $m = \dfrac{5\beta_2 - 9}{2(3-\beta_2)} = 13{\cdot}669,$

$a^2 = \dfrac{2\mu_2\beta_2}{3-\beta_2} = 1188$ [$\mu_2 = 39{\cdot}16$ in terms of unit of grouping (0·005)];

$\therefore\ a = 34{\cdot}468$ (or 0·17234 in terms of size of tetrads),

$$y_0 = \frac{N \times \Gamma(2m+2)}{a2^{2m+1}\{\Gamma(m+1)\}^2} \qquad (N = 22{,}712)$$
$$= 1412.$$

* *Tables for Statisticians and Biometricians*, Part I, p. lxiii.

Hence the equation to the best-fitting curve is

$$y = 1412\left(1 - \frac{x^2}{1188}\right)^{13\cdot669} \quad \text{(unit of grouping} = 0\cdot005).$$

Or, expressed in terms of tetrad differences,

$$y = 1412\left(1 - \frac{x^2}{0\ 17234^2}\right)^{13\ 669},$$

showing that it cuts the axis of x at the points $\pm$ 0·17234, making very "high" contact.

The best-fitting normal curve

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

gives the equation $y = 1448e^{-\frac{x^2}{78\ 32}}$ (unit of grouping = 0·005)

The accompanying figure (Fig 4) gives the symmetrical distribution of tetrad differences, with the Type IIA curve (continuous line) and normal curve (dotted line) superposed.

Applying the (P, χ^2) test for goodness of fit*, we have for the Type IIA curve, (mid-ordinates),

$$\chi^2 = S\left\{\frac{(m_r - m'_r)^2}{m_r}\right\} = 21\cdot69494$$

for 22 groups in half the symmetrical curve;

$$\therefore\ P = 0\cdot41764,$$

—a good fit, since one sample in 2·4 would give a fit as bad or worse to this curve.

For the normal curve, (areas), $\chi^2 = 46\cdot82078$,

$$\therefore\ P = 0\cdot002612,$$

—far less good a fit, although if we look at the visual pictures of the two curves in the diagram we see that the normal curve does not fall greatly behind the Type IIA curve in closeness of correspondence to the observed values of the tetrad differences. It is because of the large number of groups (22 for each half of the curve) that the numerical test of goodness of fit is so stringent and exacting in this case.

* As Professor Karl Pearson points out (11, p. 276) this test is not wholly suitable to tetrad differences, since "it is based on random selection from an infinite population, any member of which is equally likely to be drawn," but it is useful in giving us at least comparative values. For the Tables, see *Tables for Statisticians and Biometricians*, Part I, Table XII, p. 28.

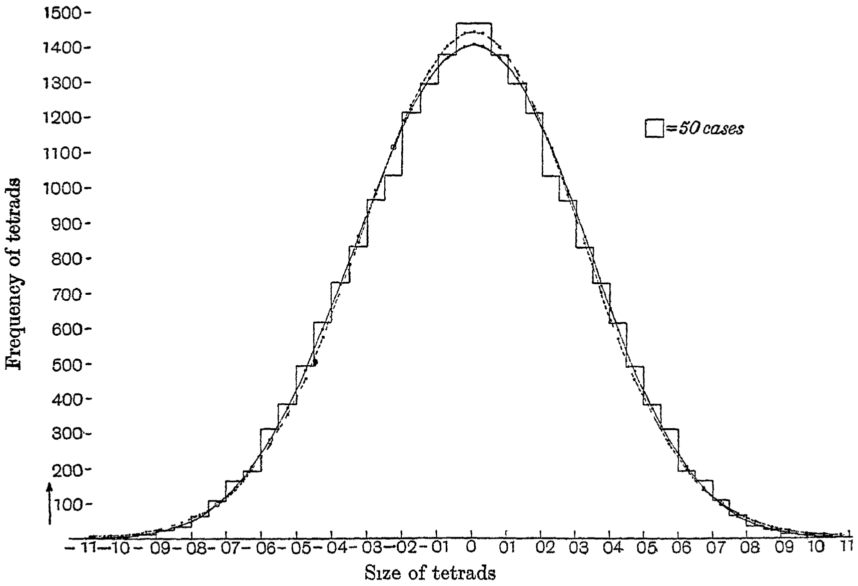


Fig 4 Symmetrical distribution of the tetrad differences

Best-fitting curve (Type IIA Pearson curve) ————————
Best-fitting probability curve - - - - - - - - - - - -

(Reproduced by permission from *Nature*, vol cxxx, p 588, October 15th, 1932)

Table IV. *Distribution of positive tetrads*

Range of tetrad	Observed frequency	Type IIA curve (mid-ordinates) $y_0=1412$	"Normal" curve (mid-ordinates) $y_0=1448$	"Normal" curve (areas)
0 000–0 005	1472	1408	1443	1441 8
0 005–0·010	1383	1376	1407	1405 6
0 010–0 015	1300	1314	1337	1335 8
0 015–0 020	1218	1225	1238	1237 0
0 020–0 025	1037	1116	1118	1117 9
0 025–0 030	968	992	984	984 7
0 030–0 035	832	861	844	843 3
0 035–0 040	731	728	706	706 4
0 040–0 045	618	599	576	576 0
0 045–0 050	493	480	457	458 0
0 050–0 055	383	373	354	355 0
0 055–0 060	311	281	268	268 2
0 060–0 065	193	205	197	197 6
0 065–0 070	163	145	141	141 8
0 070–0 075	109	99	99	99 3
0 075–0·080	65	64	67	67 7
0 080–0 085	33	40	45	45 0
0 085–0 090	24	24	29	29 2
0 090–0 095	11	14	18	18 5
0 095–0 100	8	7	11	11 4
0 100–0 105	4	4	7	6 8
Over 0 105	—	1	10	9 0
Total	11,356	11,356	11,356	11,356

The above graduations were carried out in terms of mid-ordinates in order to be able to plot the curves. For the normal curve, areas were also calculated In the case of the Type II A curve, Simpson's quadrature formula, viz.

$$\int_0^1 y\,dx = \frac{1}{6}(y_0 + 4y_{\frac{1}{2}} + y_1),$$

to change ordinate values into areas, was tried, but it made a difference of only 1·5 in the two largest groups, and less than 1·0 in all the other groups. Hence for our present purpose the correction was unnecessary*. Sheppard's corrections, for high contact, had already been applied in determining the moments of the distribution

If m be taken as 14, the nearest whole number to 13·669, the equation to the curve becomes

$$y = 1428\left(1 - \frac{x^2}{0{\cdot}17234^2}\right)^{14}.$$

For this curve $\chi^2 = 31{\cdot}71508$, therefore $P = 0{\cdot}07749$—a less good fit.

The one constant of our observed frequency-distribution which we can compare with the previously determined "theoretical" value is

$$\sigma_t = 0{\cdot}031289.$$

The excess of this over the theoretical value (0·02827) is 0 003019, i.e 1·167 (or $1\frac{1}{6}$) times the "probable error" of the latter value (0·0025866). This indicates a good correspondence of observation with theory

We may therefore conclude that, so far as this one frequency-constant is concerned, the criterion in the Theory of Two Factors satisfactorily passes the test of experience.

(10) A SUBSIDIARY TEST

The frequency-distribution of tetrads for 15 of the tests, omitting tests a, b, c, d and j of Table III, was worked out, using the same unit of grouping, 0 005. For these 4095 positive tetrads,

$$\sigma_t = 0\ 03116,$$

a result no better than that for 19 tests.

The other constants were

$$\mu_2 = 0{\cdot}000971095,\quad \mu_4 = 0{\cdot}000002486,\quad \beta_2 = 2{\cdot}6362,$$

* The high value of χ^2 for the normal curve is mainly due to the "tail" of the distribution, viz. 9 for the 22nd group "over 0 105," although this represents no more than 0·8 per 1000 distribution. Neglecting this tail, we find χ^2 for 21 groups = 37·82078, giving P = 0 01913.

giving a markedly platykurtic Type IIA curve, with equation,

$$y = 497\left(1 - \frac{x^2}{0{\cdot}1186^2}\right)^{5{\cdot}746},$$

cutting the axis of x at the tetrad values $\pm$ 0·1186.

(11) CRITICISMS MET

It has been pointed out by Professor Karl Pearson ((1), p. 247), that "if ρ_{st} represents the correlation between the sth and tth variates in an indefinitely large population which is going to be sampled, r_{st} the corresponding correlation in any particular sample, and $\bar{r}_{st}$ is the mean value of r_{st} for many samples; then $\bar{r}_{st}$ is *not* equal to ρ_{st}." Assuming normal distribution of the variates, an approximate expression of the relationship (10) is

$$\bar{r}_{st} = \rho_{st}\left[1 - \frac{1-\rho^2_{st}}{2N} + (\quad)\frac{1}{N^2} + \quad\right],$$

where N is the size of the sample

Moreover "the variation of r round ρ due to random sampling is of the order of its standard deviation, namely $1/\sqrt{N}$." Hence in writing r in the place of ρ, as we must since we do not know ρ, and as Spearman and Holzinger do in deducing their formula, terms of the order $1/N\sqrt{N}$ are neglected. If N is small this neglect may be serious. By applying our tests to 300 cases we escape the main force of this criticism.

A second criticism of a mathematical nature is that of Professor E. B. Wilson (11) and Professor H. T. H. Piaggio (12), that g is not determinate. But it has been shown by Professor Spearman (13) that this indeterminacy diminishes as the number of tests is increased, and Dr J O. Irwin (14) has proved "that g is not determinate but that its determinacy can be made as small as we please by taking a sufficient number of tests." We may fairly assume that the 19 tests employed in the present research are a large enough number to satisfy this condition.

We are still faced with the further objection of Professor Wilson that "this uniqueness (of g) is relative to the set-up, since, if we constructed artificial tests by taking linear combinations of the marks in the original tests in such a manner as to preserve the hierarchical conditions, the new g would not necessarily be the same as the old" (quoted from J. O Irwin (14), p. 363). Nevertheless we need not as psychologists be greatly disturbed by this objection.

The relation between the tetrad criterion and the "correlation between columns" (hierarchical order of correlation coefficients) criterion has recently been reconsidered in an interesting paper by Dr J. C.

Maxwell Garnett (15) One of the outcomes of his discussion seems to be the propounding of a quantity y which he conceives to be related to g as follows "g measures how much the individual tries throughout the set of tests, y measures how much good his trying does" ((15), p. 372). But he admits that this is "only a rough guess and probably a wrong one." The equation connecting the two quantities is

$$g = G + ky,$$

where G is the most probable value of g for the given mental tests of the same individual, and "k is to be determined so that the standard deviation of y is unity." The argument is directly linked up with the work of Wilson and Piaggio But since k tends towards zero as the number of tests is increased (Spearman and Irwin), the interpretation of y is not without difficulty. Dr Maxwell Garnett writes: ".. there seems no reason why y should not be *the same* for the same individual *in every set of mental tests*. If, for example, y measured the individual's 'power of concentration'—perhaps how much 'mental energy' he renders available by a unit 'effort of will'—y might be independent of his performances $q_1, \ldots, q_n$ in any particular set of tests. Then g might measure how much 'mental energy' he manifests in the set of tests, so that g would depend on y and would be common to all the tests of the set" (p. 371).

We hope that further analysis of our data may throw light on this and cognate problems, but the work must be postponed to a subsequent article. The main purpose of our research has been achieved, namely to establish Professor Spearman's Theory of Two Factors on an adequate statistical basis. We have also found that a Type IIA Pearson curve fits the distribution of tetrads for our 19 tests very closely.

Note The "theoretical" curve (Type IIA) to be expected, assuming the truth of the Two Factor Theory (Spearman), has the equation

$$y = 1563\left(1 - \frac{x^2}{0\,1557^2}\right)^{13\cdot669}.$$

See *Nature*, 1934, CXXXIII, 724.

REFERENCES

(1) PEARSON, K. and MOUL, M. The mathematics of intelligence I. The sampling errors in the theory of a generalised factor. *Biometrika*, 1927, XIX. 246.

(2) BROWN, W. The mathematical and experimental evidence for the existence of a central intellective factor (g). *Brit. Journ. Psychol* 1932, XXIII. 171, *Nature*, 1932, CXXX 588.

(3) STEPHENSON, W. Tetrad-differences for verbal subtests relative to non-verbal subtests. *J. Educ. Psych.* 1931, XXII. 334.

(4) —— Tetrad-differences for non-verbal tests. *J. Educ. Psych.* 1931, XXII. 167.

(5) LINE, W. The growth of visual perception in children *B J P. Monograph Supplement*, 1931, No. 15
(6) SPEARMAN, C Disturbers of tetrad differences *J Educ Psych.* 1930, XXI 559
(7) FORTES, M. Unpublished Thesis, University of London Library, 1930
(8) SEASHORE, C E. *Measure of Musical Talent* Columbia Graphophone Co N Y
(9) SPEARMAN, C and HOLZINGER, K. The average value for the probable error of tetrad differences *Brit Journ Psychol* 1930, XX 370
(10) SOPER, H E. On the probable error of the correlation coefficient to a second approximation *Biometrika*, 1913, IX. 105
(11) WILSON, E. B *Proc of the National Academy of Sciences*, 1928, XIV. 283
(12) PIAGGIO, H. T. H. The general factor in Spearman's theory of intelligence *Nature*, 1931, CXXVII. 56
(13) SPEARMAN, C. The theory of "two factors" and that of "sampling" *Brit. J. Educ Psych* 1931, I. 21, Note 27.
(14) IRWIN, J. O On the uniqueness of the factor *g* for general intelligence. *Brit Journ. Psychol* 1932, XXII 359.
(15) MAXWELL GARNETT, J C. Further notes on the single general factor in mental measurement. *Brit Journ. Psychol.* 1932, XXII 364

CHAPTER XIV

RECENT DEVELOPMENTS OF STATISTICAL METHOD IN PSYCHOLOGY*

[Reprinted from the autumn *Occupational Psychology*, 1938]

By GODFREY H. THOMSON

A GREAT deal of the mathematical interest in applied psychology arises from the theory of factors. The incentives to such a theory seem to me to be of two kinds, theoretical and practical; and the opinions we are likely to hold regarding it depend upon whether we are more dominated by the theoretical or the practical aspect.

(1) VOCATIONAL ADVICE

One important practical incentive is the hope that factors may be of use in vocational and educational guidance and selection. The typical form which these take, in so far as they are based upon the administration of tests, is to find the correlation coefficients of the tests with each other and with the occupation From the candidate's scores the ordinary regression equation will then give the "best" prediction of his probable ability in the occupation, in the sense that the squares of the discrepancies between the predictions and the facts, when summed over many cases, are minimised. To make such predictions more accurate, an extensive search is required for the right tests to add to the battery to increase the multiple correlation. The expense of such work, the length of time required by "follow-up" experiments, and the difficulty of getting adequate measures of the success of the candidates in their occupations, together with the great variability of the human machine, are the main obstacles to improvement in such prediction.

The practical hope of factorists has been that somehow factors would enable better predictions to be made. Now it should at once be pointed out that *mathematically* this is impossible If the use of factors turns out to improve vocational advice it will not be for any mathematical reason. For vocational or educational prediction means projecting a point given by n oblique co-ordinate axes called tests on to a vector, representing the occupation, whose direction cosines are known but

* A paper read to the Royal Society on March 24th, 1938, and published by kind permission of the Editor of the *Proceedings*

which is not in the n-space of the tests. Such estimation requires some assumption to be made about the candidate's ability along the extra dimension orthogonal to the test-space, and nothing whatever can do away with the need of such an assumption The regression method assumes that in this totally unexplored direction the candidate is average.

The use of factors, whether orthogonal or oblique, merely means referring the point in question to a new set of co-ordinate axes called factors instead of to the original test-scores, a procedure which may well have advantages of convenience for psychological thinking but cannot define the point any better, and unless care is taken may make matters worse (1), nor does the change of axes in any way facilitate the projection on to the occupation vector.

More Factors than Tests

Moreover, the task of carrying out prediction with the aid of factors as go-betweens is rendered more difficult by the circumstance that the popular systems use more factors than there are tests, so that the factors themselves have to be estimated. In addition, it is usual to estimate only what are called the common factors, throwing aside the factors which are unique to one test only If there is any guarantee that these abandoned portions of the test-variance are uncorrelated with the occupation to be predicted, then no harm is done. But the circumstances in which this guarantee can be given are precisely those circumstances in which a direct prediction without the intervention of factors can easily be made.

Maximising and Minimising Specifics

Systems which minimise the number of *common* factors have the peculiarity of thereby maximising the variance of the specific factors This maximisation of the specific variance, the part of the test-scores which is not used, must, I think, diminish the usefulness of such systems for vocational guidance. There is therefore a peculiar interest in the proposal made by M. S. Bartlett (2) to estimate factors, not on the regression principle but on the principle of minimising the squares of the specific factors summed over the tests The connection between Bartlett's estimates and the ordinary regression estimates has been deduced (3); but so far there has not been time for any practical trial of this new method. It should be noted that Bartlett accepts the number of common factors given him by others, so that his minimisation of the specifics takes place after their maximisation by Thurstone's principle.

It should also be noted that Bartlett attains his end of minimising the specifics only by making, implicitly, a different assumption regarding the ability of the candidate in traits which have not been measured and are uncorrelated with the tests. The regression method assumes that all the candidates are average in these, the Bartlett method involves assigning to each candidate different degrees of excellence in them Both are assumptions, but the former is the more likely.

A Conflict of Principle

There is, I think, need for further critical examination of the principle of minimising the number of common factors. It is defended on the grounds of parsimony. But this parsimony in the number of common factors is necessarily accompanied by prodigality in the use of specific variance. The few common factors, although they describe the correlations adequately, describe the *whole man* very inadequately, throwing away as much as possible of the information given about him by the tests. There is, in fact, a direct conflict of principle between factor methods which confine themselves to reproducing the correlations, and methods which endeavour to use all the information, excluding only what may be ascribed to sampling error.

So much for the practical side

(2) THE THEORETICAL SIDE

On the theoretical side I wish to speak briefly of three matters:

(*a*) Thurstone's conception of "Simple Structure"

(*b*) The dependence of factorial analyses on selection

(*c*) The true deduction to be drawn from the low reduced rank of correlation matrices.

There is clearly a natural desire in mankind to imagine or create, and to name, forces and powers behind the *façade* of what is observed, nor can any exception be taken to this if the hypotheses which emerge explain the phenomena as far as they go, and are a guide to further inquiry.

That the factor theory has been a guide and a spur to many investigators cannot be denied, and it is probably here that it finds its chief justification.

"Simple Structure"

The desire to find "realities" behind the phenomena appears to be strong in Thurstone. His conception of "simple structure" among factors, and his belief that when "simple structure" is achieved the

factors have a significance more than that which attaches to mere statistical coefficients, is of the greatest interest (4). His method is to break up each test into two components, of which one is orthogonal to all the other tests, while the other or communal components lie in a space of much smaller dimensions than the number of tests It is a striking fact, of which I will offer an explanation presently, that this is so generally possible. The axes of the space at which he thus arrives he rotates within that narrow common-factor-space, if necessary allowing them to become slightly oblique, until as many as possible of them are at right angles to as many as possible of the original test-vectors It is his faith that when a position can be found with a certain large number of such right angles, the axes or factors will be found to be entities acceptable to the psychologist.

It is refreshing to find so strong a belief that mathematical elegance is bound to correspond to physical or mental entities or actualities I fear, however, that the factors found from different batteries by the strict application of this mathematical principle may not correspond to one another.

It should of course be recognised that Thurstone's method is not a method of analysing *any* matrix of test-scores. It is a method of distinguishing when a set of tests is suitable for the definition of primary factors

It is clear that the attainment of "simple structure" will not be possible unless the battery of tests has been selected or purified It is desirable, therefore, to have criteria which will say rapidly from calculations on the matrix of correlations whether "simple structure" can be attained. Such criteria have recently been described, in articles discussing boundary conditions in the common-factor-space (8).

There may, however, exist different Thurstone batteries which define different and incompatible sets of factors. I think, therefore, that the factorial composition of any given test must at some stage or other be defined by the psychologist, and thereafter held invariant, for I do not think that it will naturally remain invariant If, however, such invariance is found to occur naturally, without undue forcing of the sets of tests, then this will be a very important observation. So far the evidence of this appears to me to be inadequate.

Selection among Persons

The analysis of a battery of tests into factors is very dependent upon the group of persons from whose scores the correlations are calculated, and can be changed very much by substituting a different set of persons.

Karl Pearson's fundamental formulae for changes in correlation due to selection have lately been put into very convenient matrix form by A. C. Aitken (5), and with their aid the changes in factorial analysis due to selection, which might include natural selection, can be readily followed Such selection can create or destroy factors, and change the relationship between them (6). Factors are seen as fluid descriptive mathematical coefficients, changing both with the tests used and the sample of persons, unless we take refuge in sheer definition based upon psychological judgment, which definition would have to specify the particular battery of tests, and the sample of persons, as well as the method of analysis, in order to fix any factor. This influence of selection is particularly important in vocational work, where the correlations between tests and occupations are necessarily got from selected groups, namely from men employed in each industry.

Reason for Low Rank

Lastly there is a strong tendency to be impressed by the fact that most matrices of mental correlation coefficients can be described by a comparatively small number of common factors (*plus* specifics), that is to say, that the reduced rank of such matrices tends to be low. This gives rise to a feeling that common factors must be important and actual things if they produce this remarkable effect. The fact is, however, that this tendency to a low reduced rank follows as a mathematical necessity if the causal background is structureless. Complete families of correlation coefficients will always show it if all possible samples of the causal background can be taken (7) The *less* pronounced the structure, the *stronger* the tendency to low rank in the matrix. The reason low reduced rank is so noticeable among mental correlations is that in mental measurement almost any combination of the many small causal influences can occur, whereas in measurements of height, arm-length, cranium, etc , the organs we have to measure are forced upon us. There are no such separate organs in the mind

The flux of causal influences in the background will then tend to make all matrices of mental correlations hierarchical. It is the departures from rank one which indicate the beginnings of a structure in the mind comparable with the organs of the body. The chief deduction which can be drawn from the comparatively low rank to which so many matrices of mental correlation coefficients can be reduced is, in my opinion, the conclusion that the mind of man is comparatively undifferentiated, protean and plastic, *not* that it is composed of separate faculties.

REFERENCES

(1) THOMSON, G H. Technique in Factorial Analysis. *J. Educ Psychol* 1936, XXVII. 48 and 53

(2) BARTLETT, M S. The Statistical Conception of Mental Factors. *Brit. Journ. Psychol.* 1937, XXVIII. 100

(3) THOMSON, G. H Estimating Mental Factors *Nature*, 1938, CXLI 246.

(4) THURSTONE, L. L. *The Vectors of Mind.* Chicago, 1935

(5) AITKEN, A. C. Selection from a Multivariate Normal Population *Proc. Edin. Math. Soc.* 1934, IV. 106–110

(6) THOMSON, G. H. and LEDERMANN, W. The Influence of Multivariate Selection on the Factorial Analysis of Ability. *Brit. Journ. Psychol.* 1938, XXIX 288–306.

(7) THOMSON, G. H. On Complete Families of Correlation Coefficients. *Brit. Journ Psychol.* 1935, XXVI. 63–92.

(8) LEDERMANN, W Boundary Conditions in the Factorial Analysis of Ability *Psychometrika*, 1936, I 165–174

CHAPTER XV

THE FACTORIAL ANALYSIS OF ABILITY

*THE PRESENT POSITION AND THE PROBLEMS CONFRONTING US**

[From *The British Journal of Psychology* (*General Section*), Vol XXX, Part 2, October, 1939]

By GODFREY H. THOMSON

(1) WHY DO PSYCHOLOGISTS WANT FACTORS?

I PROPOSE to set out what appear to be the main principles (some of them incompatible) of the several systems of factorial analysis which co-exist at the present day First, however, let us ask why psychologists want to have factors at all. The following reasons are given·

(*a*) If factors can be so chosen that a few of them give a good approximation to the information given by the large number of tests, there is obvious economy in their use.

(*b*) Orthogonal factors, that is, uncorrelated traits, have the advantage for scientific thought that they are independent, and they lead to simpler formulae. It should be remarked, however, that none of the human traits naturally named by naive man are uncorrelated, and that he is usually unable to realise the independence of the factors offered him by the psychologist, unable to realise for example that a man of high v (the verbal factor) is just as likely to be a man of low as of high g.

(*c*) There is a feeling that factors may be more enduring entities than the innumerable and changing tests used to find them. They come to be looked upon as the things in terms of which tests are described, although really of course it is the factors which are described in terms of tests.

(*d*) It is an easy transition to look upon the factors as actual and real. It is of the nature of man to deify or reify forces and powers behind phenomena, and we are all subject to this urge, which is, I think, a large part of the explanation of why factors are so acceptable to so many of us.

(2) THE PROBLEM OF FACTORIAL ANALYSIS

The scores of each test being set off along a line, these lines can be given directions in space at angles whose cosines are the correlations. They will then occupy a space of n dimensions if there are n tests, and the

* A contribution to the symposium presented at the Extended General Meeting of the British Psychological Society held at Reading in April, 1939.

population of persons will be represented by points in this space, congregated round the origin where the man who is average in every test is situated If standardised scores are used, the population will fall off in density equally in all directions from that point, its density contours being spheres.

The problem of factorial analysis is then to choose a set of axes or factors, preferably orthogonal or nearly so, to replace the tests as definers of the space, that is of the qualities of the persons As there are innumerable sets available, some principles must be adopted to select one as the best. "Best" may mean here best from some psychological, or from some mathematical, point of view, or possibly these may click into agreement in some system acceptable to psychologists and mathematicians alike Among the principles which have been adopted are the following*.

(i) Representing the experimental facts (within the limits of the errors present) by a smaller number of factors than tests. There is probably unanimity on this point, but not on what the smaller number of factors are to be required to do best

(ii) Reproducing as much as possible of the *whole variance* with each successive factor (Hotelling (6)).

(iii) Reproducing the *correlations* with the minimum number of common factors (Spearman (9), Thurstone).

(iv) Insisting on a *general factor g* (Spearman, Holzinger).

(v) *Rotating* the factors until they become psychologically significant (Thurstone (12, 13))

(vi) Requiring "*Simple Structure*," i.e. certain mathematical relations between factors and tests in a battery. These requirements have led to *no general factor* being found in tests reputed to be saturated with *g* (Thurstone (12, 13)).

(vii) Requiring *invariance of analysis* of the same test in different batteries, when used on equivalent samples of persons.

(viii) Using factors and loadings which are *reciprocal* for persons and tests (Burt (2, 3))

* I have placed after each, as a guide to the reader, the names of one or two of those most closely identified with the principle, but the classification is not of course perfect or complete. Kelley's name (7) will be missed. He is very important, but I have not felt able to identify him sufficiently with any of these principles Stephenson's name can accompany Spearman's wherever it appears, and he is also interested in the relationship between test factors and person factors (10), though he does not make a principle of their reciprocity as Burt does.

(3) REPRODUCING THE VARIANCE

The natural desire of the statistician (e g. Hotelling) is to lose as little information as possible He adopts therefore the second principle, using the principal axes of the ellipsoids which result from the density spheres when the test lines are pulled into orthogonality (6). It will of course require as many factors as there are tests, to reproduce all the information. But if, after a few factors, the greater part of the variance has been accounted for, the remaining factors are neglected The statistician has then arrived at the best shorthand description of *this* battery He makes no claim to have arrived at factors which have any other significance, nor that they are invariant when tests are added to or taken from the battery His axes indeed are not factors at all in the sense intended by those who use principles (v) and (vii).

There is, however, nothing to stop a worker who proceeded on these lines from then adopting principle (v), and rotating his few main factors in their own space in the hope of making them psychologically significant, but this has not in practice been done with "principal components" based on the whole variance

The first few factors found by the above method will approximate both to the total variance and also to the correlations. But it is the former they do well.

(4) REPRODUCING THE CORRELATIONS

Minimal Rank Methods

Better approximation to the correlations can be made by another method, which depends on the fact that a matrix of correlation coefficients between mental tests, when suitable numbers are inserted in its diagonal, is of low rank, within the limits of error. The reason for this is, in my belief, the complexity and plasticity of the mind, and the operation of the laws of chance. But whether this is the true reason or not is immaterial to our present argument, for the phenomenon is undoubtedly present and the result is that the correlations can be closely imitated by a small number of factors compared with the number of tests. This small number of factors does not, however, by any means give a good approximation to the whole variance but only to certain fractions called "communalities." There is therefore much information lost by using only these common factors. The other factors—specifics—are not needed to imitate the correlations of this particular battery. But they may be of importance in predicting success in some occupation. These minimal rank methods, therefore, which make much of the fact that they are parsimonious in

common factors, attain that parsimony only by contenting themselves with imitating the correlations, and giving up any attempt to reproduce all the information contained in the tests. Their second great weakness is that since the whole number of factors exceeds the number of tests, each is subject to an indeterminacy and can be estimated with less accuracy than can one of the principal components of the former method.

The Spearman School

Among those who use the minimal rank method there are two schools, that of Spearman (9), who discovered the fact of low reduced rank many years ago, and that of Thurstone (12). The method of the former school is to proceed by purifying a battery of intelligence tests until one general factor g is sufficient to account for the correlations. The g-saturations of the tests in this battery are then known. In more complex batteries this g is always taken out first. When a second factor is suspected, a battery is built up to measure it (by the residuals left after the removal of g), a battery which contains only these two. And so the process goes on step by step, a complex matrix of correlations being looked upon as made up by the superposition on top of the g-hierarchy, of sub-hierarchies each due to a new group factor. Invariance of analysis can be gained (principle (vii)) by retaining for any test that g-saturation which it had in the best "g-battery," and so on with the other factors. No rotation is necessary.

Holzinger's Bifactor Method (4) is the same in principle, although it proceeds on more wholesale lines.

The Thurstone School

In contrast with the step by step Spearman method, the Thurstone method is a simultaneous one. No attempt is made to reduce any battery by purification to rank one. Instead, a large battery is analysed simultaneously into a number of common factors—plus the specifics characteristic of both schools—by a method which involves first of all guessing the communalities and then taking at each stage the centroid of the residues. The resulting factors are devoid of psychological significance and have to be rotated in a search for positions where they will have such significance.

Here Thurstone introduces the idea of "Simple Structure," a mathematically defined position of the axes (12). It is his belief that if in any battery "Simple Structure" can be attained the factors will then be recognisable as psychological entities. "Simple Structure" is marked by the absence of negative saturations and *the presence of many zero saturations*, for it is laid down that there must be at least one zero saturation in each test, and several zero saturations in each factor. This excludes a

general factor, and Thurstone naturally therefore does not find one (13), even when he analyses tests said to be highly saturated with g. Just as naturally Holzinger and Harman (5), analysing the same data, do find a g, its presence in their pattern being, as they frankly say, "due to our hypothesis of its existence."

Alexander (1) has carried out an analysis using Thurstone's way of arriving at a set of minimal rank factors, but not his hypothesis of "Simple Structure." Instead, Alexander at this point passes over to the Spearman plan, and adopts for some of his tests analyses already known. This throws into relief the fact that the essential difference between Thurstone and Spearman is in the former's use of "Simple Structure." Spearman has no need, in his step by step method, to rotate his factors. But if he did use Thurstone's first analysis (as Alexander has done) he could then rotate to positions corresponding to the results of the step by step procedure. If, however, a battery of tests was put together from Spearman analyses, and offered for alternative analysis via "Simple Structure," it would be necessary, to give a chance of agreement, to include several tests with nearly zero saturation in g, otherwise *either* "Simple Structure" would not be attainable, *or* the two methods would be fore-ordained to disagree.

(5) RECIPROCITY OF TESTS AND PERSONS

The last principle (reciprocity of persons and tests) has been put forward in its rigorous form by Burt only (2, 3), although Stephenson (10) has proposed to check analyses of tests and of persons by one another. Burt holds that those factors are the proper ones to use of which it can be said that the same results are arrived at whether we analyse tests or persons—the factors and loadings of the one analysis being the loadings and factors of the other. This end, however, Burt attains only by removing from the operation of the principle a factor consisting of the average performance of each person in all tests, and by working with unstandardised scores and covariances instead of correlations. Actually he removes his "average" factor not by analysis but by selection of cases equal in average score*. On a matrix of scores which has somehow been centred both ways his reciprocity principle however does work. Such a matrix has only $n - 1$ dimensions (one factor, the average, having been removed) and Burt uses the principal axes of its ellipsoids. His

* I would like to take this opportunity of withdrawing lines 33 to 35 on page 290 of my book *The Factorial Analysis of Human Ability* (11), for I now see that since Burt chose his sub-sample of persons to be not only equal in average to one another, but equal to the average of all, those lines are irrelevant.

factors are therefore free from "indeterminacy." He does not, however, rotate them but interprets them as they are, with their numerous negative loadings, a procedure which has passed muster in the analysis of temperament to which alone so far he has applied it, but would arouse opposition were it applied to the analysis of more ordinary mental test-scores. His use of raw scores just as they stand is certainly wrong: but if a suitable set of units could be discovered—I have suggested one based on the sampling theory (11)—the practice of analysing covariances instead of correlations would have much to commend it.

(6) VARIOUS OTHER MATTERS

There remain several matters which I have space only to mention. For a long time workers were content to analyse batteries of tests into factors, and did not proceed to the practical estimation of these factors in individuals Now that this is beginning to be done, a realisation is growing of how uncertain such estimates are, as a result of the mathematical weakness of postulating in minimal rank systems more factors than tests. It is also coming to be realised that to estimate factors, and from these to estimate proficiency in an occupation, is a roundabout procedure when the latter can be estimated direct.

Finally, much has been done to illustrate the influence on factors of selection among the persons, and also of training and maturing. In connection with selection it has been made clear that factors are bound to change, emerge, and disappear with the changing sample, and therefore also with the generations of man, and Price (8) has indicated how homogamy can influence the factorial make-up of successive generations Factors are statistical coefficients, changing with the sample and the conditions and dependent upon stated assumptions: but with defined conditions and assumptions they are most useful as descriptive terms.

(7) A SUMMING UP

In summing up I remind myself first of all that neither Thurstone nor any keen follower of his has been present at this symposium to defend the idea of "Simple Structure." Let me therefore for one paragraph hold a temporary brief on his behalf.

I recently heard Professor Dirac uphold, before the Royal Society of Edinburgh, the thesis that when a mathematical physicist finds a mathematical solution or theorem which is particularly beautiful—and every mathematician knows what mathematical beauty is—he can have considerable confidence that it will prove to correspond to something real in

physical nature. Something of this same faith seems to me to lie behind Professor Thurstone's trust in the coincidence of "Simple Structure" in the matrix of factor loadings with psychological significance in the factors thus defined. This is of course not all. The conditions laid down in "Simple Structure" are laid down in order to remove the embarrassing number of degrees of freedom which permit the factor axes to be rotated like a Catherine wheel. They clamp the axes at one point (unless very unusual conditions prevail in the battery) if indeed that point can be reached, which is not necessarily the case in every battery. If it cannot actually be reached it can be approached as nearly as possible And Thurstone also gives, in pp. 71 and 72 of *Primary Mental Abilities*, and perhaps elsewhere, what one might call his common-sense reasons for thinking it plausible that numerous zeros will occur among the loadings. Moreover, we must remember that he has embarked on a campaign of forming augmented sub-batteries round tests chosen from his original 57, to obtain better definition of his factors, and has obtained what one must admit to be rather encouraging results in his study of the perceptual factor (14).

The disagreement between Thurstone and Spearman as to the general factor g is perhaps the most important point of controversy, as it is certainly one of the most easily grasped by the non-mathematical psychologist and one of the most disquieting to him. Professor Spearman has defended the general factor which will always be associated with his own name; but I think (crossing the court now to speak on his side of the case) that he might have done so more fundamentally than he has. For the objections he has raised to Thurstone's work might conceivably all be overcome. A longer time could be given to the tests. Product-moment instead of tetrachoric correlations could be calculated. Error could be reduced by greater care, and by using larger and better samples of subjects. Yet still Thurstone might be able—indeed, with a properly chosen battery, I think he almost certainly would be able—to attain "Simple Structure" among his loadings, and the complaint that a dismembered g was merely being submerged in a sea of error would no longer be justified, at any rate as far as the submergence was concerned. Surely the real defence of g is simply that it has proved useful. It still remains to be seen whether Thurstone's primary factors will prove equally useful. A jury in this country would probably to-day give a verdict to Spearman; a jury in another place might give a verdict to Thurstone. The larger jury of the future will, I think, decide by noting which system has proved more useful in the hands of the practising psychologist.

I myself lean at the moment more towards Spearman's g and his

later group factors than I do to Thurstone's, since they seem to me more in accord with the ideas of my own Sampling Theory. On that theory g is as it were the whole mind, and tests are part of g, not g part of the tests And were that mind entirely undifferentiated, structureless, g would be the only factor we need. As the complexity of the mind, or the complexity of the upper brain, is organised (partly by the maturing of hereditary bonds, mainly I fancy by education and life) and integrated into "pools," "clusters," call them what you will, so additional factors, additional descriptive coefficients, are needed. It seems to me at present wise to retain one coefficient to express the general depth starting from which the integration, the deepening into subpools, has gone on. But I am not sure, and think the better course is to await further papers from workers in Thurstone's school. I think it possible that there will come a stage when practical progress will be most facilitated by a firm voluntary agreement among psychologists that certain well-known tests have such and such a factorial composition, are not to be changed, and may be used to fix some of the factorial axes in a battery The sample of persons would also have to be defined and different factorial compositions might be given for children and adults, for literates and illiterates, etc., etc.

REFERENCES

(1) ALEXANDER, W P. Intelligence, concrete and abstract *Brit Journ Psychol. Monogr Suppl* 1935, No. 19

(2) BURT, C Correlation between persons *Brit Journ Psychol* 1937, XXVIII 59–96.

(3) —— The analysis of temperament. *Brit Journ. med Psychol* 1938, XVII 158–188.

(4) HOLZINGER, K. J *Preliminary Reports of Spearman-Holzinger Unitary Trait Study.* No 5 *Introduction to Bifactor Theory* Chicago, 1935

(5) HOLZINGER, K J and HARMAN, H H Comparison of two factorial analyses. *Psychometrika*, 1938, III 45–60.

(6) HOTELLING, H Analysis of a complex of statistical variables into principal components *J Educ Psychol* 1933, XXIV. 417–441 and 498–520.

(7) KELLEY, T. L *Essential Traits of Mental Life* Harvard Univ Press, 1935.

(8) PRICE, B Homogamy and the intercorrelations of capacity traits *Ann Eugen., Lond* 1936, VII. 22–27

(9) SPEARMAN, C. *The Abilities of Man* London Macmillan, 1927

(10) STEPHENSON, W. The foundations of psychometry four factor systems *Psychometrika*, 1936, I. 195–209

(11) THOMSON, GODFREY H. *The Factorial Analysis of Human Ability* London Univ. Press, and Houghton Mifflin, Boston, 1939.

(12) THURSTONE, L. L. *The Vectors of Mind.* Univ Chicago Press, 1935.

(13) —— Primary mental abilities. *Psychometric Monogr.* No. 1. Univ. Chicago Press, 1938

(14) —— The perceptual factor. *Psychometrika*, 1938, III. 1–18.

APPENDIX I

TABLES

1. Fechner's Fundamental Table.
2. Urban's Tables for the Constant Process.
3. Table of Muller-Urban Weights
4. Reciprocals of pq, where $p + q = 1$.
5. Rich's Checking Table for the Constant Process.

1. *Fechner's Fundamental Table*

p	γ	p	γ	p	γ
0 50	0 0000	0 67	0 3111	0 84	0 7031
0 51	0 0177	0 68	0 3307	0 85	0 7329
0 52	0 0355	0 69	0 3506	0 86	0 7639
0 53	0 0532	0 70	0 3708	0 87	0 7965
0 54	0 0710	0 71	0 3913	0 88	0 8308
0 55	0 0888	0 72	0 4121	0 89	0 8673
0 56	0 1067	0 73	0 4333	0 90	0 9062
0 57	0 1247	0 74	0 4549	0 91	0 9480
0 58	0 1427	0 75	0 4769	0 92	0 9935
0 59	0 1609	0 76	0 4994	0 93	1 0435
0·60	0 1792	0 77	0 5224	0 94	1 0993
0 61	0 1975	0 78	0 5460	0 95	1 1630
0·62	0 2160	0 79	0 5702	0 96	1 2380
0 63	0 2346	0 80	0 5951	0 97	1 3300
0 64	0 2535	0 81	0 6208	0 98	1 4520
0 65	0 2725	0 82	0 6473	0 99	1 6450
0 66	0 2916	0 83	0 6747	1 00	∞

If p is $< 0{\cdot}5$, look in the Table not for p but for $1 - p$, and take γ negative. Thus the γ for $p = 0{\cdot}25$ is $- 0{\cdot}4769$.

This Table is less generally useful than Sheppard's Tables of the Probability Integral, Tables I and II of Pearson's *Tables for Statisticians and Biometricians* (Cambridge University Press). Table I there differs from this only by a factor $\sqrt{2}$: but gives many more values and to more decimal places.

2 *Urban's Tables for the Constant Process.*

From *Archiv f d ges. Psychol.* 1912, XXIV 240—241*.

p	W	γW	$2 W$	$2^2 W$	$2 \gamma W$	$3 W$	$3^2 W$	$3 \gamma W$	$4 W$	$4^2 W$	$4 \gamma W$
0 50	1 0000	0 0000	2 0000	4 0000	0 0000	3 0000	9 0000	0 0000	4 0000	16 0000	0 0000
0 51	0 9998	0 0177	1 9996	3 9991	0 0354	2 9993	8 9980	0 0531	3 9991	15 9965	0 0708
0 52	0 9991	0 0355	1 9982	3 9963	0 0709	2 9972	8 9917	0 1064	3 9963	15 9853	0 1419
0·53	0 9980	0 0531	1 9959	3 9918	0 1062	2 9938	8 9816	0 1593	3 9918	15 9672	0 2124
0 54	0 9964	0 0707	1 9928	3 9855	0 1415	2 9891	8 9674	0 2122	3 9855	15 9421	0 2830
0 55	0 9943	0 0883	1 9886	3 9772	0 1766	2 9829	8 9487	0 2649	3 9772	15 9088	0 3532
0 56	0 9918	0 1058	1 9836	3 9671	0 2116	2 9753	8 9260	0 3175	3 9671	15 8685	0 4233
0 57	0 9888	0 1233	1 9776	3 9551	0 2466	2 9663	8 8990	0 3699	3 9551	15 8205	0 4932
0 58	0 9853	0 1406	1 9706	3 9413	0 2812	2 9560	8 8679	0 4218	3 9413	15 7651	0 5624
0 59	0 9814	0 1579	1 9627	3 9254	0 3158	2 9441	8 8322	0 4737	3 9254	15 7018	0 6316
0 60	0 9768	0 1750	1 9537	3 9074	0 3501	2 9306	8 7916	0 5252	3 9074	15 6296	0 7002
0 61	0 9720	0 1920	1 9440	3 8881	0 3839	2 9161	8 7482	0 5759	3 8881	15 5523	0 7679
0 62	0 9666	0 2088	1 9332	3 8663	0 4176	2 8997	8 6992	0 6263	3 8663	15 4653	0 8351
0 63	0 9607	0 2254	1 9214	3 8429	0 4508	2 8822	8 6465	0 6762	3 8429	15 3715	0 9015
0 64	0 9542	0 2419	1 9084	3 8168	0 4838	2 8626	8 5878	0 7257	3 8168	15 2672	0 9676
0 65	0 9473	0 2581	1 8945	3 7890	0 5163	2 8418	8 5253	0 7744	3 7890	15 1562	1 0325
0 66	0 9398	0 2741	1 8797	3 7594	0 5481	2 8196	8 4586	0 8222	3 7594	15 0376	1 0962
0 67	0 9317	0 2899	1 8634	3 7268	0 5797	2 7951	8 3853	0 8696	3 7268	14 9072	1 1594
0 68	0 9232	0 3053	1 8464	3 6929	0 6106	2 7697	8 3090	0 9159	3 6929	14 7715	1 2212
0 69	0 9140	0 3205	1 8280	3 6561	0 6409	2 7421	8 2262	0 9614	3 6561	14 6243	1 2818
0 70	0 9043	0 3353	1 8085	3 6170	0 6706	2 7128	8 1383	1 0059	3 6170	14 4682	1 3412
0 71	0 8939	0 3498	1 7878	3 5755	0 6996	2 6816	8 0449	1 0493	3 5755	14 3021	1 3991
0 72	0 8830	0 3639	1 7659	3 5318	0 7277	2 6489	7 9466	1 0916	3 5318	14 1274	1 4555
0 73	0 8713	0 3775	1 7426	3 4852	0 7551	2 6139	7 8417	1 1326	3 4852	13 9408	1 5101
0 74	0 8590	0 3908	1 7180	3 4360	0 7815	2 5770	7 7310	1 1723	3 4360	13 7440	1 5630
0 75	0 8460	0 4035	1 6921	3 3842	0 8070	2 5381	7 6144	1 2104	3 3842	13 5366	1 6139
0 76	0 8323	0 4157	1 6646	3 3293	0 8313	2 4970	7 4909	1 2470	3 3293	13 3171	1 6626
0 77	0 8179	0 4273	1 6357	3 2714	0 8545	2 4536	7 3607	1 2818	3 2714	13 0858	1 7090
0·78	0 8025	0 4382	1 6051	3 2102	0 8764	2 4076	7 2229	1 3146	3 2102	12 8406	1 7527
0 79	0 7865	0 4484	1 5729	3 1459	0 8969	2 3594	7 0782	1 3453	3 1459	12 5835	1 7938
0 80	0 7695	0 4579	1 5390	3 0780	0 9159	2 3085	6 9255	1 3738	3 0780	12 3120	1 8317
0 81	0 7515	0 4665	1 5031	3 0061	0 9331	2 2546	6 7638	1 3996	3 0061	12 0245	1 8662
0 82	0 7327	0 4743	1 4653	2 9307	0 9485	2 1980	6 5940	1 4228	2 9307	11 7227	1 8970
0·83	0 7129	0 4810	1 4257	2 8515	0 9619	2 1386	6 4158	1 4429	2 8515	11 4059	1 9239
0 84	0 6921	0 4866	1 3842	2 7683	0 9732	2 0762	6 2287	1 4598	2 7683	11 0733	1 9464
0 85	0 6697	0 4908	1·3394	2 6788	0 9816	2 0091	6 0273	1 4725	2 6788	10 7152	1 9633
0 86	0 6463	0 4937	1 2927	2 5853	0 9875	1 9390	5 8170	1 4812	2 5853	10 3413	1 9749
0 87	0 6215	0 4950	1 2430	2 4860	0 9900	1 8645	5·5935	1 4851	2 4860	9 9440	1 9801
0 88	0 5953	0 4946	1 1907	2 3813	0 9892	1 7860	5 3580	1 4838	2 3813	9 5253	1 9784
0 89	0 5673	0 4920	1·1346	2 2692	0 9840	1 7019	5 1056	1 4760	2 2692	9 0766	1 9680
0 90	0 5376	0 4871	1 0751	2 1502	0 9743	1 6126	4 8380	1 4614	2 1502	8 6008	1 9485
0 91	0 5059	0 4796	1 0118	2 0236	0 9592	1 5177	4 5531	1 4388	2 0236	8 0944	1 9184
0 92	0·4718	0 4687	0 9435	1 8871	0 9374	1 4153	4 2459	1 4061	1 8871	7 5483	1 8748
0 93	0 4351	0 4540	0 8702	1 7403	0 9080	1 3052	3 9157	1 3620	1 7403	6 9613	1 8160
0·94	0 3954	0 4346	0 7907	1 5814	0 8692	1 1861	3 5582	1 3039	1 5814	6 3258	1 7385
0 95	0 3519	0 4093	0 7038	1 4076	0 8185	1 0557	3 1671	1 2278	1 4076	5 6304	1 6370
0 96	0 3036	0 3759	0 6073	1·2146	0 7518	0 9109	2 7328	1 1277	1 2146	4 8582	1 5036
0 97	0 2469	0 3282	0 4936	0 9871	0 6564	0 7403	2 2210	0 9847	0 9871	3 9485	1 3129
0 98	0·1881	0 2732	0 3762	0 7525	0 5463	0 5644	1 6931	0 8195	0 7525	3 0099	1 0926
0 99	0·1127	0 1854	0 2254	0 4508	0 3708	0 3381	1 0142	0 5561	0 4508	1 8030	0 7415

* With three corrections, two of which were made by Urban in the *Praxis der Konstanzmethode*, and the third by Rich in *Amer Journ. Psychol* 1918, XXIX 121. The last was also rediscovered by the *Cambridge Univ Press* proof reader

2. *Urban's Tables for the Constant Process* (*contd.*).

From *Archiv f d ges Psychol* 1912, xxiv 240—241

p	$5\,W$	$5^2\,W$	$5\gamma\,W$	$6\,W$	$6^2\,W$	$6\gamma\,W$	$7\,W$	$7^2\,W$	$7\gamma\,W$
0 50	5 0000	25 0000	0 0000	6 0000	36 0000	0 0000	7 0000	49 0000	0 0000
0 51	4 9989	24 9945	0 0885	5 9987	35 9921	0 1062	6 9985	48 9892	0 1239
0 52	4 9954	24 9770	0 1773	5 9945	35 9669	0 2128	6 9936	48 9549	0 2483
0 53	4 9898	24 9488	0 2655	5 9877	35 9262	0 3185	6 9856	48 8996	0 3716
0 54	4 9819	24 9095	0 3537	5 9783	35 8697	0 4251	6 9747	48 8266	0 4960
0 55	4 9715	24 8575	0 4415	5 9658	35 7948	0 5298	6 9601	48 7207	0 6181
0 56	4 9589	24 7945	0 5291	5 9507	35 7041	0 6349	6 9425	48 5972	0 7408
0 57	4 9439	24 7195	0 6165	5 9327	35 5961	0 7398	6 9215	48 4502	0 8631
0 58	4 9266	24 6330	0 7030	5 9119	35 4715	0 8436	6 8972	48 2807	0 9842
0 59	4 9068	24 5340	0 7895	5 8882	35 3290	0 9474	6 8695	48 0866	1 1053
0 60	4 8842	24 4212	0 8753	5 8611	35 1666	1 0503	6 8380	47 8656	1 2254
0 61	4 8601	24 3005	0 9599	5 8321	34 9927	1·1518	6 8041	47 6290	1 3438
0 62	4 8329	24 1645	1 0439	5 7995	34 7969	1 2527	6 7661	47 3624	1 4615
0 63	4 8036	24 0180	1 1269	5 7643	34 5859	1 3523	6 7250	47 0753	1 5777
0 64	4 7710	23 8550	1 2094	5 7252	34 3512	1 4513	6 6794	46 7558	1 6932
0 65	4 7363	23 6815	1 2906	5 6836	34 1014	1 5488	6 6308	46 4157	1 8069
0 66	4 6992	23 4962	1 3703	5 6391	33 8346	1 6444	6 5790	46 0526	1 9184
0 67	4 6585	23 2925	1 4493	5 5902	33 5412	1 7391	6 5219	45 6533	2 0290
0 68	4 6161	23 0805	1 5265	5 5393	33 2359	1 8319	6 4625	45 2378	2 1372
0 69	4 5701	22 8505	1 6023	5 4841	32 9047	1 9227	6 3981	44 7870	2 2432
0 70	4 5213	22 6065	1 6765	5 4256	32 5534	2 0118	6 3298	44 3087	2 3471
0 71	4 4694	22 3470	1 7489	5 3633	32 1797	2 0987	6 2572	43 8001	2 4484
0 72	4 4148	22 0740	1 8193	5 2978	31 7866	2 1832	6 1807	43 2650	2 5471
0 73	4 3565	21 7825	1 8877	5 2278	31 3668	2 2652	6 0991	42 6937	2 6427
0 74	4 2950	21 4750	1 9538	5 1540	30 9240	2 3446	6 0130	42 0910	2 7353
0 75	4 2302	21 1510	2 0174	5 0762	30 4574	2 4209	5 9223	41 4560	2 8243
0 76	4 1616	20 8080	2 0783	4 9939	29 9635	2 4940	5 8262	40 7837	2 9096
0 77	4 0893	20 4465	2 1363	4 9072	29 4430	2 5635	5 7250	40 0751	2 9908
0 78	4 0127	20 0635	2 1909	4 8152	28 8914	2 6291	5 6178	39 3245	3 0673
0 79	3 9324	19 6618	2 2422	4 7188	28 3129	2 6907	5 5053	38 5370	3 1391
0 80	3 8475	19 2375	2 2896	4 6170	27 7020	2 7476	5 3865	37 7055	3 2055
0 81	3 7576	18 7882	2 3327	4 5092	27 0551	2 7993	5 2607	36 8250	3 2658
0 82	3 6634	18 3168	2 3713	4 3960	26 3761	2 8455	5 1287	35 9008	3 3198
0 83	3 5644	17 8218	2 4049	4 2772	25 6631	2 8858	4 9901	34 9306	3 3668
0 84	3 4604	17 3020	2 4330	4 1525	24 9149	2 9196	4 8447	33 9119	3 4062
0 85	3 3485	16 7425	2 4541	4 0182	24 1092	2 9449	4 6879	32 8153	3 4358
0 86	3 2316	16 1582	2 4687	3 8780	23 2679	2 9624	4 5243	31 6702	3 4561
0 87	3 1075	15 5375	2 4751	3 7290	22 3740	2 9701	4 3505	30 4535	3 4652
0 88	2 9766	14 8832	2 4730	3 5720	21 4319	2 9676	4 1673	29 1712	3 4622
0 89	2 8364	14 1822	2 4600	3 4037	20 4222	2 9521	3 9710	27 7972	3 4441
0 90	2 6878	13 4388	2 4356	3 2253	19 3518	2 9228	3 7628	26 3400	3 4099
0 91	2 5295	12 6475	2 3980	3 0354	18 2124	2 8776	3 5413	24 7891	3 3572
0 92	2 3588	11 7942	2 3435	2 8306	16 9837	2 8122	3 3024	23 1167	3 2809
0 93	2 1754	10 8770	2 2700	2 6105	15 6629	2 7240	3 0456	21 3189	3 1780
0 94	1 9768	9 8840	2 1731	2 3722	14 2330	2 6077	2 7675	19 3726	3 0423
0 95	1·7595	8 7975	2 0463	2 1114	12 6684	2 4556	2 4633	17 2431	2 8648
0 96	1 5182	7 5910	1 8795	1 8218	10 9310	2 2554	2 1255	14 8784	2 6313
0 97	1 2339	6 1695	1 6411	1 4807	8 8841	1 9693	1 7275	12 0922	2 2975
0 98	0 9406	4 7030	1 3658	1 1287	6 7723	1 6389	1 3168	9 2179	1 9121
0 99	0 5634	2 7172	0 9269	0 6761	3 9568	1 1123	0 7888	5 4218	1 2976

Note, 1924 A fourth error discussed by H. H Long has been corrected, v *Amer. Journ Psychol* 1922, xxxiii p 303.

3. *Table of Müller-Urban Weights**.

p	W	p	W	p	W
0·50	1·000	0·67	0·932	0·84	0·694
0·51	1·000	0·68	0·923	0·85	0·670
0·52	0·999	0·69	0·914	0·86	0·646
0·53	0·998	0·70	0·904	0·87	0·621
0·54	0·996	0·71	0·894	0·88	0·595
0·55	0·995	0·72	0·883	0·89	0·567
0·56	0·992	0·73	0·871	0·90	0·538
0·57	0·989	0·74	0·859	0·91	0·506
0·58	0·985	0·75	0·846	0·92	0·472
0·59	0·981	0·76	0·832	0·93	0·435
0·60	0·977	0·77	0·818	0·94	0·396
0·61	0·972	0·78	0·803	0·95	0·352
0·62	0·967	0·79	0·787	0·96	0·304
0·63	0·960	0·80	0·770	0·97	0·249
0·64	0·954	0·81	0·752	0·98	0·187
0·65	0·947	0·82	0·733	0·99	0·112
0·66	0·940	0·83	0·713	1·00	0·000

The weight of a p which is less than 0·5 is the same as the weight of a p which exceeds 0·5 by the same amount. Thus the weights of $p = 0{\cdot}25$ and of $p = 0{\cdot}75$ are both alike, $= 0{\cdot}846$.

* The table is quoted from F. M. Urban, "The Method of Constant Stimuli and its Generalisations," *Psychological Review*, 1910, XVII. p. 253. See also "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen," by the same author, *Archiv f. d. ges. Psychologie*, XV. and XVI.

4. *Reciprocals of pq, where $p + q = 1$.*

p or q	$\frac{1}{pq}$	p or q	$\frac{1}{pq}$	p or q	$\frac{1}{pq}$
·50	4·0	·67	4·5	·84	7·5
·51	4·0	·68	4·6	·85	7·9
·52	4·0	·69	4·7	·86	8·3
·53	4·0	·70	4·8	·87	8·8
·54	4·0	·71	4·9	·88	9·4
·55	4·1	·72	5·0	·89	10·2
·56	4·1	·73	5·1	·90	11·1
·57	4·1	·74	5·2	·91	12·2
·58	4·1	·75	5·3	·92	13·6
·59	4·1	·76	5·4	·93	15·4
·60	4·2	·77	5·6	·94	17·7
·61	4·2	·78	5·8	·95	21·0
·62	4·3	·79	6·0	·96	26·0
·63	4·3	·80	6·2	·97	34·4
·64	4·4	·81	6·5	·98	51·0
·65	4·4	·82	6·8	·99	101
·66	4·5	·83	7·1	1·00	∞

5. *Rich's Checking Table for the Constant Method.*

Published from the Laboratory of Cornell University, in *Amer Journ Psychol.* 1918, XXIX 120.

An example will best explain the use of this Table. Consider the example worked on p. 73. If we form for each line the totals of the five quantities

$$W + \gamma W + sW + s^2W + s\gamma W$$

we get the results

s	p	Totals
−6	00	0000
−5	10	13 2371
−4	14	9 8835
−3	40	7 1880
−2	65	2 5836
0	·80	1 2274
2	87	5 8355
4	·96	8 2559
6	1 00	0 0000
		48 2110

These totals however will all be found in Rich's Table entered with the proper s (Rich's x) and p, and can thus be checked. The grand total 48·2110 ought to agree with the sum of the last row of the table (4) on p 73, i.e. with

$$4{\cdot}8026 + 0{\cdot}4311 - 7{\cdot}6406 + 43{\cdot}7049 + 6{\cdot}9130.$$

Rich's Checking Table.

p	$x=1$	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$	$x=7$
01	– 0327	2327	7235	1 4396	2 2810	3 4479	4 8403
02	0179	4973	1 3529	2 5847	4 1927	6 1770	8 5375
03	0843	7430	1 8953	3 5414	5 6810	8 3142	11 4409
·04	·1590	9978	2 4437	4 4969	7 1574	10 4251	14 3003
·05	2371	1 2355	2 9376	5 3436	8 4533	12 2668	16 7842
06	3170	1 4637	3 4012	6 1295	9 6485	13 9583	19 0586
07	3973	1 6836	3 8400	6 8667	10 7635	15 5305	21 1676
08	4780	1 8963	4 2582	7 5637	11 8126	17 0052	23 1413
09	·5585	2 1025	4 6583	8 2259	12 8053	18 3965	24 9995
10	6386	2 3015	5 0397	8 8530	13 7415	19 7048	26 7434
11	7179	2 4951	5 4068	9 4531	14 6339	20 9491	28 3994
12	·7967	2 6835	5 7609	10 0289	15 4875	22 1370	29 9770
13	8745	2 8655	6 0994	10 5764	16 2964	23 2594	31 4653
14	9515	3 0431	6 4274	11 1043	17 0737	24 3361	32 8910
15	1 0275	3 2155	6 7428	11 6096	17 8158	25 3614	34 2463
16	1 1031	3 3848	7 0506	12 1007	18 5349	26 3533	35 5559
17	1 1767	3 5472	7 3434	12 5654	19 2132	27 2864	36 7858
18	1 2495	3 7059	7 6276	13 0148	19 8673	28 1850	37 9681
19	1 3215	3 8611	7 9038	13 4494	20 4981	29 0500	39 1049
20	1 3927	4 0127	8 1718	13 8699	21 1070	29 8830	40 1981
21	1 4627	4 1600	8 4304	14 2737	21 6901	30 6791	41 2413
22	1 5311	4 3032	8 6802	14 6624	22 2496	31 4418	42 2393
23	1 5991	4 4432	8 9231	15 0388	22 7901	32 1773	43 1999
·24	1 6655	4 5792	9 1575	15 4004	23 3079	32 8800	44 1169
25	1 7310	4 7118	9 3846	15 7494	23 8063	33 5552	44 9965
26	1 7954	4 8407	9 6039	16 0852	24 2844	34 2016	45 8369
27	1 8589	4 9665	9 8168	16 4097	24 7451	34 8232	46 6439
28	1 9212	5 0891	10 0230	16 7228	25 1886	35 4203	47 4177
29	1 9821	5 2078	10 2213	17 0226	25 6116	35 9884	48 1530
30	2 0423	5 3239	10 4142	17 3130	26 0203	36 5362	48 8604
31	2 1010	5 4367	10 6004	17 5921	26 4118	37 0596	49 5354
32	2 1590	5 5466	10 7807	17 8611	26 7880	37 5612	50 1810
33	2 2153	5 6523	10 9526	18 1164	27 1435	38 0341	50 7880
34	2 2712	5 7567	11 1217	18 3665	27 4908	38 4950	51 3789
·35	2 3257	5 8564	11 2819	18 6019	27 8164	38 9254	51 9288
36	2 3788	5 9537	11 4370	18 8287	28 1289	39 3374	52 4543
·37	2 4313	6 0488	11 5878	19 0482	28 4300	39 7332	52 9579
38	2 4822	6 1397	11 7304	19 2543	28 7113	40 1015	53 4248
·39	2 5320	6 2282	11·8684	19 4525	28 9807	40 4530	53 8693
·40	2 5804	6 3128	11 9988	19 6386	29 2319	40 7792	54 2800
·41	2 6284	6 3958	12 1261	19 8191	29 4748	41 0933	54 6743
·42	2·6747	6 4754	12 2468	19 9887	29 7013	41 3845	55 0384
·43	2 7198	6 5516	12 3609	20 1479	29 9124	41 6545	55 3741
·44	2 7638	6 6251	12 4698	20 2983	30 1103	41 9059	55 6849
45	2 8063	6 6952	12 5727	20 4388	30 2935	42 1368	55 9687
·46	2 8478	6 7625	12 6700	20 5703	30 4634	42 3486	56 2310
·47	2 8878	6·8264	12 7610	20 6915	30 6180	42 5403	56 4585
·48	2 9263	6·8872	12 8461	20 8033	30 7587	42 7122	56 6638
·49	2 9640	6·9454	12·9263	20 9069	30 8870	42 8667	56 8459
·50	3 0000	7·0000	13·0000	21·0000	31·0000	43 0000	57 0000

Rich's Checking Table.

p	$x=1$	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$	$x=7$
51	3 0348	7 0516	13 0679	21 0839	31 0994	43 1145	57 1291
52	3 0683	7 1000	13 1299	21 1581	31 1843	43 2088	57 2314
53	3 1002	7 1450	13 1858	21 2225	31 2552	43 2835	57 3079
54	3 1306	7 1869	13 2358	21 2777	31 3122	43 3402	57 3644
55	3 1595	7 2250	13 2791	21 3218	31 3531	43 3730	57 3815
56	3 1870	7 2599	13 3164	21 3565	31 3801	43 3873	57 3781
57	3 2130	7 2914	13 3473	21 3809	31 3920	43 3807	57 3469
58	3 2371	7 3190	13 3716	21 3947	31 3885	43 3529	57 2880
59	3 2600	7 3432	13 3893	21 3981	31 3696	43 3039	57 2007
60	3 2804	7 3630	13 3992	21 3890	31 3325	43 2298	57 0808
61	3 3000	7 3800	13 4042	21 3723	31 2845	43 1406	56 9409
62	3 3174	7 3925	13 4006	21 3421	31 2167	43 0245	56 7654
63	3 3329	7 4012	13 3910	21 3020	31 1346	42 8886	56 5641
64	3 3464	7 4051	13 3722	21 2477	31 0315	42 7238	56 3245
65	3 3581	7 4052	13 3469	21 1831	30 9138	42 5392	56 0588
66	3 3676	7 4011	13 3143	21 1071	30 7796	42 3320	55 7639
67	3 3749	7 3915	13 2716	21 0150	30 6219	42 0921	55 4258
68	3 3802	7 3784	13 2231	20 9141	30 4516	41 8356	55 0660
69	3 3830	7 3595	13 1642	20 7967	30 2574	41 5460	54 6628
70	3 3835	7 3357	13 0966	20 6660	30 0439	41 2304	54 2252
71	3 3813	7 3066	13 0195	20 5204	29 8090	40 8854	53 7494
72	3 3768	7 2723	12 9340	20 3616	29 5550	40 5145	53 2397
73	3 3689	7 2317	12 8370	20 1849	29 2755	40 1086	52 6843
·74	3 3586	7 1853	12 7301	19 9928	28 9736	39 6724	52 0891
75	3 3450	7 1328	12 6124	19 7842	28 6481	39 2040	51 4521
76	3 3283	7 0732	12·4829	19 5570	28 2959	38 6994	50 7675
77	3 3083	7 0068	12·3413	19 3114	27 9173	38 1589	50 0361
78	3 2839	6 9324	12 1858	19 0442	27 5078	37 5764	49 2503
79	3 2563	6 8506	12 0178	18 7581	27 0713	36 9573	48 4163
·80	3 2243	6 7603	11 8352	18 4491	26 6020	36 2940	47 5249
·81	3 1875	6 6603	11 6360	18 1148	26 0965	35 5816	46 5695
82	3 1467	6 5515	11 4218	17 7574	25 5585	34 8246	45 5563
83	3 1007	6 4330	11 1912	17 3752	24 9850	34 0200	44 4814
84	3 0495	6 3044	10 9434	16 9667	24 3741	33 1657	43 3415
85	2 9907	6 1603	10 6694	16 5178	23 7056	32 2328	42 0995
86	2 9263	6 0055	10 3772	16 0415	22 9985	31 2483	40 7906
87	2 8545	5 8355	10 0596	15 5266	22 2366	30 1896	39 3857
88	2 7751	5 6511	9 7177	14 9749	21 4227	29 0614	37 8906
89	2 6859	5 4471	9 3428	14 3731	20 5379	27 8373	36 2716
90	2 5870	5 2243	8 9367	13 7242	19 5869	26 5246	34 5374
91	2 4769	4 9801	8 4951	13 0219	18 5605	25 1109	32 6731
92	2 3528	4 7085	8 0078	12 2507	17 4370	23 5670	30 6405
93	2 2133	4 4076	7 4720	11 4067	16 2115	21 8865	28 4316
·94	2 0554	4 0713	6 8782	10 4757	14 8639	20 0429	26 0124
95	1 8743	3 6911	6 2118	9 4362	13 3645	17 9966	23 3324
·96	1 6626	3 2532	5 4509	8 2559	11 6682	15 6877	20 3147
97	1 3971	2 7122	4 5211	6 8236	9 6196	12 9092	16 6923
98	1 1107	2 1363	3 5383	5 3163	7 4707	10 0012	12 9081
99	·7089	1 3451	2 2065	3 2934	4 5056	6 0433	7 8063

Rich's Checking Table

p	$x=-7$	$x=-6$	$x=-5$	$x=-4$	$x=-3$	$x=-2$	$x=-1$	$x=0$
01	5 8579	4 3203	3 0080	2 0210	1 1595	5235	1127	− 0727
02	9 7281	7 1974	5 0431	3 2649	1 8631	8375	1881	− 0851
03	12 5809	9 2914	6 4954	4 1930	2 3841	1 0686	2469	− 0813
04	15 3119	11 2923	7 8800	5 0749	2 8773	1 2868	3036	− 0723
05	17 5872	12 9552	9 0269	5 8024	3 2818	1 4649	3519	− 0574
06	19 6082	14 4293	10 0411	6 4437	3 6368	1 6207	3954	− 0392
07	21 4324	15 7575	10 9527	7 0181	3 9536	1 7592	·4351	− 0189
08	23 0983	16 9684	11 7820	7 5391	4 2398	1 8841	4718	0031
09	24 6313	18 0809	12 5423	8 0155	4 5005	1 9973	·5059	0263
10	26 0376	19 0998	13 2371	8 4496	4 7373	2 0999	5376	0505
11	27 3456	20 0459	13 8811	8 8507	4 9550	2 1939	5673	0753
12	28 5668	20 9282	14 4803	9 2231	5 1565	2 2805	5953	1007
13	29 6947	21 7416	15 0316	9 5646	5 3406	2 3595	6215	1265
14	30 7546	22 5049	15 5479	9 8835	5 5118	2 4327	6463	1526
15	31 7421	23 2148	16 0270	10 1786	5 6696	2 4999	6697	1789
16	32 6789	23 8875	16 4801	10 4569	5 8178	2 5628	6921	2055
17	33 5392	24 5036	16 8942	10 7102	5 9520	2 6196	7129	2319
18	34 3503	25 0840	17 2831	10 9474	6 0772	2 6723	7327	2584
19	35 1151	25 6302	17 6483	11 1696	6 1938	2 7211	7515	2850
20	35 8361	26 1442	17 9912	11 3773	6 3024	2 7665	·7695	3116
21	36 5089	26 6229	18 3097	11 5695	6 4022	2 8080	7865	3381
22	37 1383	27 0696	18 6060	11 7474	6 4942	2 8458	8025	3643
23	37 7315	27 4899	18 8841	11 9140	6 5795	2 8808	8179	3906
24	38 2837	27 8802	19 1413	12 0670	6 6575	2 9126	8323	4166
25	38·8005	28 2446	19 3807	12 2088	6 7292	2 9416	8460	4425
·26	39 2815	28 5828	19 6020	12 3392	6 7945	2 9677	8590	4682
27	39 7311	28 8980	19 8075	12 4595	6 8542	2 9915	8713	4938
28	40 1505	29 1911	19 9976	12 5702	6·9084	3 0127	8830	5191
29	40 5354	29 4592	20 1706	12 6698	6 9567	3 0314	8939	5441
30	40 8950	29 7086	20 3307	12 7614	7 0004	3 0481	9043	5690
31	41 2256	29 9368	20 4762	12 8435	7 0390	3 0625	9140	5935
32	41 5304	30 1464	20 6088	12 9177	7 0731	3 0750	9232	6179
·33	41·8022	30 3319	20 7251	12 9816	7 1016	3 0849	9317	6418
34	42 0577	30 5056	20 8330	13 0401	7 1269	3 0935	9398	6657
35	42·2810	30 6558	20 9250	13 0889	7 1471	3 1000	9473	6892
36	42 4819	30 7896	21 0057	13 1303	7 1632	3·1045	·9542	7123
37	42 6633	30 9092	21 0766	13 1654	7 1758	3 1076	·9607	7353
38	42 8156	31 0079	21 1333	13 1919	7 1836	3 1085	9666	7578
39	42 9487	31 0924	21 1803	13 2121	7 1880	3 1080	9720	7800
40	43 0548	31 1576	21 2141	13 2242	7 1880	3 1056	9768	8018
41	43 1459	31 2117	21 2402	13 2315	7 1853	3 1020	9814	8235
·42	43 2124	31 2479	21 2541	13 2309	7 1784	3·0966	9853	8447
43	43 2573	31 2687	21 2576	13 2241	7 1681	3 0896	9888	8655
44	43 2815	31 2743	21 2507	13 2107	7 1542	3 0811	9918	8860
45	43 2847	31 2648	21 2335	13·1908	7 1367	3 0712	·9943	9060
·46	43·2736	31 2422	21 2070	13 1653	7·1162	3 0599	·9964	9257
·47	43 2305	31 2019	21 1694	13 1327	7 0920	3 0470	·9980	9449
·48	43 1732	31 1488	21 1225	13 0945	7 0645	3 0326	9991	9636
·49	43 0967	31 0817	21 0662	13 0503	7 0339	3 0170	9998	9821
·50	43 0000	31 0000	21 0000	13 0000	7·0000	3 0000	1 0000	1 0000

Rich's Checking Table

p	$x=-7$	$x=-6$	$x=-5$	$x=-4$	$x=-3$	$x=-2$	$x=-1$	$x=0$
51	42 8843	30 9047	20 9246	12 9441	6 9631	2 9816	9998	1 0175
52	42 7476	30 7942	20 8389	12 8817	6 9227	2 9618	9991	1 0346
53	42 5935	30 6711	20 7446	12 8141	6 8796	2 9408	9980	1 0511
54	42 4230	30 5334	20 6410	12 7407	6 8332	2 9183	9964	1 0671
55	42 2251	30 3818	20 5271	12 6610	6 7835	2 8946	9943	1 0826
56	42 0115	30 2161	20 4041	12 5757	6 7308	2 8695	9918	1 0976
57	41 7777	30 0357	20 2712	12 4843	6 6749	2 8430	9888	1 1121
58	41 5252	29 8419	20 1293	12 3873	6 6160	2 8154	9853	1 1259
59	41 2511	29 6327	19 9770	12 2841	6 5537	2 7862	9814	1 1393
60	40 9540	29 4070	19 8135	12 1738	6 4876	2 7554	9768	1 1518
61	40 6451	29 1728	19 6445	12 0603	6 4202	2 7242	9720	1 1640
62	40 3102	28 9201	19 4631	11 9393	6 3486	2 6909	9666	1 1754
63	39 9587	28 6554	19 2736	11 8132	6 2742	2 6568	9607	1 1861
64	39 5793	28 3708	19 0707	11 6789	6 1956	2 6207	9542	1 1961
65	39 1834	28 0744	18 8600	11 5401	6 1145	2 5836	9473	1 2054
66	38 7691	27 7650	18 6406	11 3959	6 0307	2 5455	9398	1 2139
67	38 3240	27 4335	18 4063	11 2426	5 9422	2 5053	9317	1 2216
68	37 8666	27 0932	18 1664	11 0859	5 8519	2 4644	9232	1 2285
69	37 3802	26 7324	17 9126	10 9209	5 7572	2 4217	9140	1 2345
70	36 8714	26 3556	17 6483	10 7496	5 6592	2 3775	9043	1 2396
71	36 3382	25 9614	17 3724	10 5712	5 5577	2 3318	8939	1 2437
72	35 7841	25 5525	17 0868	10 3870	5 4530	2 2851	8830	1 2469
73	35 2007	25 1226	16 7871	10 1943	5 3440	2 2363	8713	1 2488
74	34 5925	24 6752	16 4760	9 9948	5 2315	2 1863	8590	1 2498
·75	33 9589	24 2098	16 1529	9 7880	5 1154	2 1346	8460	1 2495
76	33 2959	23 7236	15 8161	9 5732	4 9949	2 0814	8323	1 2480
77	32 6045	23 2175	15 4661	9 3506	4 8705	2 0264	8179	1 2452
·78	31 8801	22 6878	15 1006	9 1184	4 7414	1 9694	8025	1 2407
79	31 1275	22 1383	14 7221	8 8787	4 6084	1 9110	7865	1 2349
80	30 3409	21 5648	14 3278	8 6297	4 4706	1 8505	7695	1 2274
81	29 5165	20 9646	13 9159	8 3702	4 3276	1 7879	7515	1 2180
82	28 6593	20 3416	13 4891	8 1020	4 1802	1 7239	7327	1 2070
83	27 7676	19 6940	13 0464	7 8244	4·0282	1 6578	7129	1 1939
84	26 8397	19 0215	12 5873	7 5373	3 8714	1 5896	6921	1 1787
85	25 8521	18 3066	12 1004	7 2336	3 7062	1 5183	6697	1 1605
86	24 8298	17 5675	11 5979	6 9211	3 5368	1 4451	6463	1 1400
87	23 7543	16 7914	11 0714	6 5944	3 3604	1 3695	6215	1 1165
88	22 6316	15 9822	10 5235	6 2555	3 1781	1 2913	5953	1 0899
89	21 4414	15 1257	9 9451	5 8987	2 9870	1 2099	5673	1 0593
90	20 1920	14 2284	9·3401	5 5268	2 7887	1 1255	5376	1 0247
91	18 8761	13 2849	8 7055	5 1379	2 5821	1 0381	5059	9855
92	17 4739	12 2814	8 0324	4 7269	2 3650	9467	4718	9405
93	15 9844	11 2175	7 3207	4 2941	2 1376	8512	4351	8891
94	14 3928	10 0831	6 5641	3 8359	1 8982	7515	3954	8300
·95	12 6762	8 8626	5 7529	3 3470	1 6448	6465	3519	7612
·96	10 8011	7 5333	4 8728	2 8195	1 3737	5350	3036	6795
97	8 6423	6 0092	3 8696	2 2236	1 0711	4122	·2469	·5751
98	6 4503	4 4660	2 8579	1·6261	7705	2913	1881	·4613
99	3 6335	2 4665	1 5250	·9088	·4181	·1527	1127	2981

APPENDIX II

A LIST OF DEFINITE INTEGRALS OF FREQUENT OCCURRENCE IN PROBABILITY WORK

$$A = \int_0^\infty z e^{-z^2}\,dz = \tfrac{1}{2}, \quad \int_{-\infty}^{\infty} = \text{zero}.$$

$$B = \int_0^\infty e^{-z^2}\,dz = \frac{\sqrt{\pi}}{2}, \quad \int_{-\infty}^{\infty} = \sqrt{\pi}.$$

$$C = \int_0^\infty z^2 e^{-z^2}\,dz = \frac{\sqrt{\pi}}{4}, \quad \int_{-\infty}^{\infty} = \frac{\sqrt{\pi}}{2}.$$

$$D = \int_0^\infty z^3 e^{-z^2}\,dz = \tfrac{1}{2}, \quad \int_{-\infty}^{\infty} = \text{zero}.$$

$$E = \int_0^\infty z^n e^{-z^2}\,dz = \frac{1\,.\,3\,.\,5\,\ldots\,(n-1)}{2\,.\,2\,.\,2\,..\,\;2}\,\frac{\sqrt{\pi}}{2}\ (n \text{ even})$$

$$\text{or } \frac{1\,.\,2\;\;4\,.\,6\,\ldots\,(n-1)}{2\,.\,2\,.\,2\,.\,2\;\,..\,2}\ (n \text{ odd}).$$

$$F = \int_0^{\frac{\pi}{2}} \frac{d\theta}{P\cos^2\theta + Q\sin^2\theta} = \frac{\pi}{2\sqrt{(PQ)}}, \quad \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{\pi}{\sqrt{(PQ)}}.$$

$$G = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{d\theta}{P\cos^2\theta - 2R\cos\theta\sin\theta + Q\sin^2\theta} = \frac{\pi}{\sqrt{(PQ-R^2)}},$$

$$\text{but } \int_0^{\frac{\pi}{2}} \text{ is } \textit{not } \frac{G}{2}.$$

$$H = \int_{-\infty}^{\infty} e^{-(P+2Rz+Qz^2)}\,dz = \frac{\sqrt{\pi}}{\sqrt{Q}}\,e^{-\frac{PQ-R^2}{Q}}, \text{ but } \int_0^\infty \text{ is } \textit{not } \frac{H}{2}.$$

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(Px^2+2Rxy+Qy^2)}\,dx\,dy = \frac{\pi}{\sqrt{(PQ-R^2)}},$$

$$\text{if } PQ > R^2, \text{ but } \int_0^\infty\int_0^\infty \text{ is } \textit{not } \frac{J}{4}.$$

$$K = \int_{-\infty}^{\infty} e^{-(P+2Rz+Qz^2)}\,z\,dz = -\frac{R\sqrt{\pi}}{Q^{\frac{3}{2}}}\,e^{-\frac{PQ-R^2}{Q}}, \text{ but } \int_0^\infty \text{ is } \textit{not } \frac{K}{2}.$$

$$L = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(Px^2+2Rxy+Qy^2)}\, xy\, dx\, dy = -\frac{\pi R}{2\,(PQ-R^2)^{\frac{3}{2}}} \quad \text{if } PQ > R^2,$$

$$\text{but } \int_0^{\infty}\int_0^{\infty} \text{ is } \textit{not } \frac{L}{4}.$$

$$M = \int_{-\infty}^{\infty} e^{-(P+2Rz+Qz^2)}\, z^2 dz = \frac{\sqrt{\pi}\,(Q+2R^2)}{2Q^{\frac{5}{2}}}\, e^{-\frac{PQ-R^2}{Q}}.$$

$$N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(Px^2+2Rxy+Qy^2)}\, x^2 dx\, dy = \frac{\pi Q}{2\,(PQ-R^2)^{\frac{3}{2}}}, \text{ if } PQ > R^2$$

$$O = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(Px^2+2Rxy+Qy^2)}\, \frac{y}{x}\, dx\, dy = -\frac{\pi R}{Q\sqrt{(PQ-R^2)}}, \text{ if } PQ > R^2$$

NOTE In the above integrals P and Q do *not*, as often, represent constants whose sum is unity R is *not* a correlation coefficient, although it is closely connected therewith

INDEX

CAMBRIDGE. PRINTED BY W. LEWIS, M.A., AT THE UNIVERSITY PRESS